Heidi Repo

Structural studies on lysosomal proteins

Research Program in Structural Biology and Biophysics
Institute of Biotechnology, University of Helsinki

Genetics, Department of Biosciences,
Faculty of Biological and environmental sciences, University of Helsinki

Integrative Life Science Doctoral Program
and
National Doctoral Programme in Informational and Structural Biology

ACADEMIC DISSERTATION

To be presented for public examination with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki in the auditorium 2 of Info Centre Korona, Viikinkaari 11, Helsinki, on June 26th 2014 at 12 o´clock noon.

Helsinki 2014

Supervised by

Adjunct professor Pirkko Heikinheimo
Department of Biochemistry
Faculty of Mathematics and
Natural Sciences
University of Turku, Finland

Professor Adrian Goldman
Department of Biosciences
Faculty of Biological and
Environmental Sciences
University of Helsinki, Finland
and
Faculty of Biological Sciences
University of Leeds, UK

Thesis committee members

Adjunct professor Pirkko Heikinheimo

Professor Adrian Goldman

Adjunct professor Eeva-Liisa Eskelinen
Department of Biosciences
Faculty of Biological and
Environmental Sciences
University of Helsinki, Finland

Reviewed by

Adjunct professor Tiina Salminen
Department of Biosciences
Åbo Akademi University
Turku, Finland

Adjunct professor Aija Kyttälä
National Institute for Health
and Welfare
Public Health Genomics
Helsinki, Finland

Opponent

Adjunct professor Tuomo Glumoff
Faculty of Biochemistry and
Molecular Medicine
University of Oulu, Finland

Custos

Professor Juha Partanen
Genetics
Department of Biosciences
Faculty of Biological and
Environmental Sciences
University of Helsinki, Finland

ISBN 978-952-10-9971-7 (paperback)
ISBN 978-952-10-9972-4 (PDF, http://ethesis.helsinki.fi/)
ISSN 1799-7372

Helsinki University Printing House

Helsinki 2014

Rakkaan isäni muistolle

In loving memory of my father

Abbreviations

acyltransferase	acyl coenzyme A isopenicillin N acyltransferase
AGA	N(4)-(β-*N*-acetylglucosaminyl)-L-asparaginase
CHO cells	Chinese hamster ovary cells
DNJ	deoxynojirimycin
EE	early endosome
EET	enzyme enhancement therapy
ER	endoplasmic reticulum
ERAD	endoplasmic reticulum associated degradation
ERT	enzyme replacement therapy
GAA	lysosomal α-glucosidase
GALNS	*N*-acetylgalactosamine-6-sulphatase
GH31	glycoside-hydrolase family 31
GLA	α-galactosidase A
Glc	glucose
GlcNAc	*N*-acetylglucosamine residue
GlcNAc phosphotransferase	*N*-acetylglucosamine-1-phosphotransferase
LAMP-1	lysosome-associated membrane protein-1
LAMP-2	lysosome-associated membrane protein-2
LAMP-3	lysosome-associated membrane protein-3
LE	late endosome
LIMP-2	lysosome integral membrane protein 2
LSD	lysosomal storage disorder
Man	mannose
MAN2B1	lysosomal α-mannosidase
MGA	maltase-glucoamylase
M6P	mannose-6-phosphate
N-glycan	asparagine linked glycosylation
Ntn	N-terminal nucleophile
Ntn-hydrolase	N-terminal nucleophile aminohydrolase
PLB	phospholipase B
PLBD1	phospholipase B-like protein 1
bPLBD1	bovine phospholipase B-like protein 1
hPLPD1	human phospholipase B-like protein 1

PLBD2	phospholipase B-like protein 2
hPLBD2	human phospholipase B-like protein 2
mPLBD2	mouse phospholipase B-like protein 2
RE	recycling endosome
SL	secretory lysosome
UPS	ubiquitin proteasome system

Abbreviations of amino acids

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
W	Trp	Tryptophan
V	Val	Valine
Y	Tyr	Tyrosine

List of original publications

This thesis is based on following publications:

I Bruckmann C*, Repo H*, Kuokkanen E, Xhaard H and Heikinheimo P (2012). Systematic Structure–Activity Study on Potential Chaperone Lead Compounds for Acid α-Glucosidase. *ChemMedChem* 7: 1943–1953

II Repo H, Kuokkanen E, Oksanen E, Goldman A, Heikinheimo P (2014). Is phospholipase B-like protein an amidase? *Proteins: Structure, Function, and Bioinformatics* 82: 300-311.

III Pokharel K*, Repo H* and Heikinheimo P. What makes a protein lysosomal? *Manuscript.*

*These authors contributed equally.

The publications are referred to in the text by their Roman numerals (I-III). Thesis contains previously unpublished material.

Abstract

Lysosomes are membrane enclosed acidic cell organelles found ubiquitously in higher eukaryotes. The lysosomal lumen contains more than 60 soluble lysosomal hydrolases, which degrade and recycle cellular macromolecules. Mutations in genes encoding lysosomal or lysosome related proteins result in over 50 different lysosomal storage disorders (LSDs) affecting 1 out of every 7700 newborn children. For instance, the first described LSD, Pompe disease, is caused by a mutation that impairs the function of lysosomal α-glucosidase (GAA) and that results in lysosomal accumulation of glycogen. In this study, several lysosomal proteins were studied via a variety of techniques to increase the knowledge of lysosomal function and correspondingly, the lysosome associated diseases.

In order to better understand its function, the previously not well characterised phospholipase B-like protein 1 (PLBD1) was purified from bovine kidneys. It was crystallised and the structure solved by X-ray crystallography to a 1.9 Å resolution. The structure showed that PLBD1 is a member of the N-terminal nucleophile aminohydrolases superfamily. This would imply that PLBD1 is not an esterase as the name suggests, but an amidase. The finding that the hydrophobic tail of the potential phospholipid substrate does not fit into the acyl binding cavity also argues against phosphoesterase function.

As a first step in the protein transport pathway to lysosomes mannose-6-phosphate-tag is added to lysosomal proteins. This is initiated by *N*-acetylglucosamine-1-phosphotransferase (GlcNAc phosphotransferase), which requires a recognition signal on the folded surface of the lysosomal proteins. In this study, the conservation of the signal in four lysosomal proteins was analysed. The phosphorylated N-glycosylation sites and the lysine residues on the GlcNAc phosphotransferase recognition site are well conserved at the sequence level in orthologous proteins, but not necessarily in the protein family. Based on surface analysis of PLBD1 and comparison to the paralogous PLBD2, the most likely recognition site for the GlcNAc phosphotransferase for the PLBD1 could be suggested.

LSD associated mutations affect protein function through several mechanism. Several disease-associated missense mutations disturb the protein fold. A general analysis of four enzymes associated with LSD showed that the disease-associated missense mutations are not equally distributed among the 20 amino acids. Glycine, arginine and proline are clearly over-represented among the mutations compared

to their abundance in protein sequences. The hydrophobic amino acids tend to be under-represented among disease-associated mutations. The amino acids where mutation frequently involves a disease have unique properties that contribute to the protein structure in a way that cannot be compensated by other amino acids.

Enzyme enhancement therapy with chemical chaperones is a novel treatment for LSDs and has shown potential also for Pompe disease. In this study, the stabilisation capacity of potential chemical chaperones for GAA were tested. Most of the compounds stabilised rhGAA against thermal unfolding and some stabilised even better than would be expected from their binding affinity. In addition, the compounds were modelled to the active site of a GAA structural model and based on this three factors to be considered in chemical chaperone design were defined. Firstly, the ligand size can vary, but the four OH-groups in the ligand are critical in orienting the molecule and making the binding specific. Last but most importantly, a positive charge and its location determine the strength of binding to GAA.

This thesis with its structural studies of lysosomal proteins provides molecular understanding of lysosomal protein biology, which is critical for full understanding lysosome function and its involvement in diseases.

1 Introduction

In each of the three domains of life, bacteria, archaea and eukaryota, the cell is the basic unit of life. Eukaryotic cells are distinct in that in all of them, different cellular functions are arranged to distinct compartments called cell organelles. In multicellular organisms, the diversity of cell organelles facilitates specialisation of the cells to different roles. Most cells share a critical set of organelles for their basic functions. In animal cells, the lysosome, the main degradative compartment of the cell, is one of them.

Lysosomes were discovered in the 1950's when Christian de Duve used tissue fractionation to characterise hepatic glucose-6-phosphatase (de Duve *et al*., 1955). He named the 'acid precipitable cell structures' as lysosomes, after the Greek words for a 'digestive body' (de Duve *et al*., 1955; de Duve, 2005). de Duve received a Nobel prize of Physiology and Medicine for his findings in 1974. The first lysosomal storage disorder was described by H. G. Hers in 1963, when he reported the absence of lysosomal α-glucosidase in Pompe disease (Hers, 1963).

In the case of the lysosome, the separation of the organelle's functions from the rest of the cell has several advantages. The membrane of a lysosome limits the hydrolytic activities of the lysosomal enzymes to the lumenal side, and so protects other cell components from non-specific hydrolysis. The separate compartment also allows the maintenance of the acidic environment in the lysosomal lumen, which is beneficial for the hydrolytic processes. In addition, since transport of molecules to lysosomes is highly controlled, cells can regulate what is hydrolysed and when the hydrolysis takes place. Thus, degradation can be used both for recycling and regulation purposes.

The lysosomal lumen is now known to contain more than 60 soluble lysosomal hydrolases. They are a versatile group of mostly unrelated proteins, of which some are quite well characterised, such as the lysosomal α-glucosidase. Nonetheless, there are still poorly understood proteins, such as the the phospholipase B-like protein 1, described in this study. It is also very likely that not all of the lysosomal proteins with hydrolytic activity have yet been identified.

Structural biology addresses the questions on molecular and structural basis of biology, through the three dimensional structures of biological macromolecules. The information gained in structural biology forms a strong basis for most protein research. X-ray crystallography is the standard method for solving three dimensional structures of proteins at atomic resolution.

Today, the structures of approximately two thirds of the more than 60 soluble lysosomal hydrolases are known and the atomic coordinates are available in the protein databank (www.rcsb.org; Berman *et al*., 2000). The analysis of lysosomal protein structures benefits research both in biological and medical problems on lysosomes. For instance, protein structures give insight to lysosomal protein processing (Coulombe *et al*., 1996), low pH activation (Heikinheimo *et al*., 2003) and transport to lysosomes (Steet *et al*., 2005; Warner *et al*., 2002). Important clues of the active site location and the catalytic mechanism of several hydrolases, such as N(4)-(β-*N*-acetylglucosaminyl)-L-asparaginase (AGA) (Oinonen *et al*., 1995) and arylsulfatase A (Lukatela *et al*., 1998), are derived from the analysis of the protein structures. In addition, protein structures have provided insight to the molecular basis for several LSDs such as α-mannosidosis (Kuokkanen *et al*., 2011) and Schindler and Kanzaki diseases (Clark and Garman, 2009). In addition, the development of therapeutic approaches for LSDs, such as the enzyme enhancement therapy (EET), has used structural information as a basis for understanding the mode of action (Parenti, 2009).

2 Review of the literature

2.1 Role of lysosome in cells

Lysosomes are membrane enclosed cell organelles found ubiquitously in higher eukaryotes. Their number differs greatly between cells. In addition, the morphology and size of individual lysosomes varies even inside the same cell (Figure 1). Lysosomes are characterised by low pH and by the presence of several hydrolases with optimum activity at acidic pH (Schröder *et al*., 2010). The hydrolases are the key for the main function of the lysosome, which is the breakdown and recycling of cellular macromolecules.

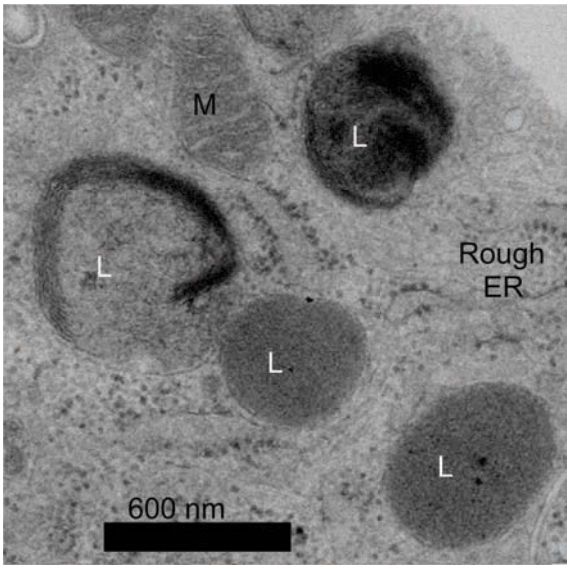


Figure 1. Electron micrograph of mouse embryonic fibroblasts. *Lysosomes appear as electron dense globular bodies in electron micrographs. Depending on their content they can be multilaminar, have a fairly homogenous appearance, or have distinct vesicles inside them. ER stands for endoplasmic reticulum M for Mitochondrion and L for lysosomes. The electron micrograph was kindly provided by Dr. Eeva-Liisa Eskelinen, Department of Biosciences, University of Helsinki, Finland.*

The material that is degraded in lysosomes is diverse and includes proteins, sugars, macromolecular assemblies, worn out cell organelles and microbes. Several pathways deliver material to lysosomes. The main pathways for internalised extracellular and cell surface material are the endocytic pathways (Straus, 1958; de Duve and Wattiaux, 1966; Luzio *et al*., 2009) and for the intracellular material the autophagy pathways (Hruban *et al*., 1963; de Duve and

Wattiaux, 1966; Eskelinen and Saftig, 2009). Equally important are the salvage pathways that work in the opposite direction. They transport degraded material out of the lysosome for re-use in biosynthesis or for energy production.

The lysosome and the multitude of its transport pathways form the greater lysosomal system (Walkley, 2009). The greater lysosomal system is highly complex and extremely dynamic (Figure 2). During oxidative stress, for example, proteins from mitochondria can be selectively removed to vesicles and delivered to lysosomes for degradation (Soubannier *et al.*, 2012); and after damage or during starvation, entire mitochondria can be degraded by autophagocytosis (Kim *et al.*, 2007). The biogenesis of lysosomes is also finely regulated. For instance,

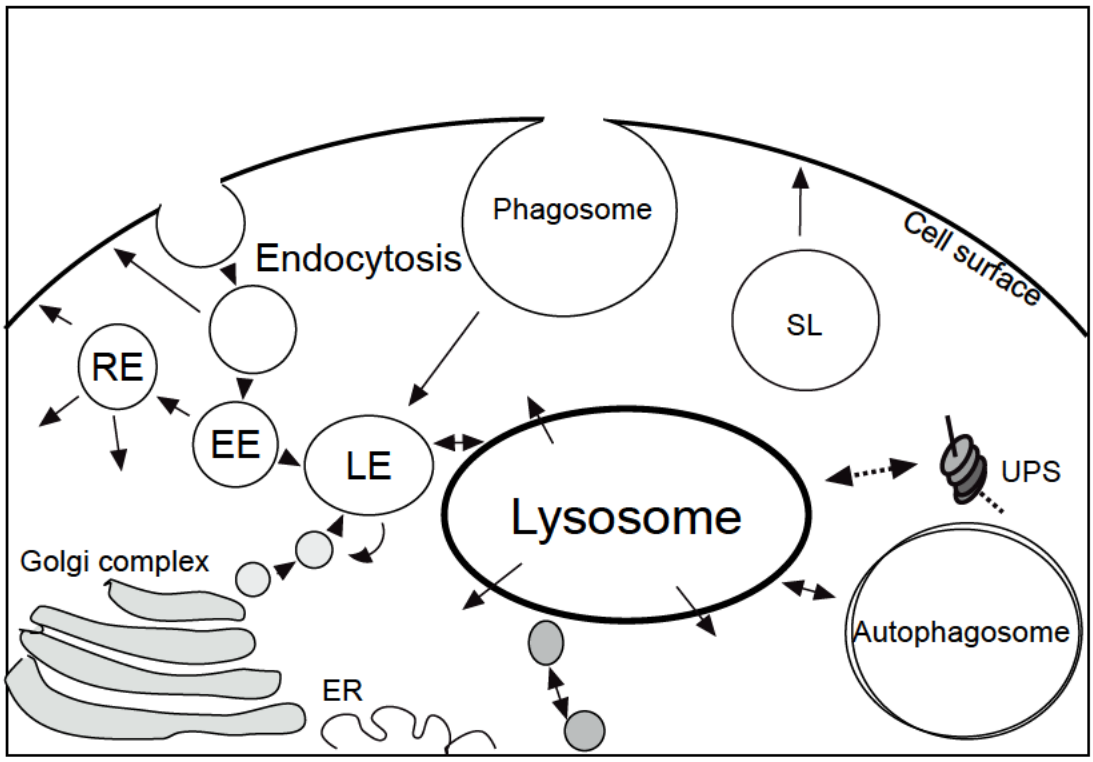


Figure 2. Schematic drawing of the main components of the greater lysosomal system. *Vesicles transport endocytosed material to early endosomes (EE), which also receive ubiquitinated membrane proteins destined for degradation. From the EE, material is either recycled rapidly or sorted to recycling endosomes (RE) or to late endosomes (LE). In specialised cells, large particles are internalised by a form of endocytosis called phagocytosis and delivered to LEs. LEs deliver material for degradation either by fusing or by "kiss and run" with lysosomes. Secretory lysosomes (SL) are utilised in a few cell types. The internalised cytosolic content of autophagosomes is degraded after fusing with lysosomes. The ubiquitin proteasome system (UPS) can coordinate proteolysis with lysosomes (Korolchuk et al., 2010; Zhao et al., 2007). Newly synthesised molecules from the Golgi apparatus are transported to LEs and from there sorted to lysosomes or recycled back to the Golgi. Salvage routes deliver degraded material from lysosomes to different parts of the cell or out of the cell.*

most genes related to the formation and function of lysosomes have coordinated expression, which is regulated by the transcription factor EB (Sardiello *et al.*, 2009).

Since they were discovered, lysosomes have been recognised to be involved in many functions in the cell. For instance, the nervous system develops and maintains neurons by trafficking and degrading neurotrophic factors and their receptors through the endocytic pathway (Glerup *et al*., 2013). In thyroids, endocytosis is used to process the prohormone thyroglobulin (Friedrichs *et al*., 2003). Secretory lysosomes and phagosomes function mostly in the cells of the immune system. Cytotoxic T-cells use their secretory lysosomes, the lytic granules, for target cell killing (Russell and Ley, 2002). Neutrophils utilise both phagocytosis and secretory lysosomes for pathogen killing (Lee *et al*., 2003) as do macrophages although by different mechanisms. In osteoblasts, secretory lysosomes are involved in bone resorption (Baron *et al*., 1985). Phagocytosis is used also by non immune system cells such as fibroblasts to ingest apoptotic cells and thus participate in cell turn over (Elliott and Ravichandran, 2010).

Thus, in addition to their basic degradative function, lysosomes have important maintenance function in multicellular organisms. Lysosomes are involved at least in bone and tissue homeostasis and remodelling, cholesterol homeostasis, immunodefence, cell signalling and hormone and growth factor regulation.

2.2. Soluble lysosomal protein synthesis and transport

2.2.1 Synthesis and post translational modifications

Lysosomal proteins are synthesised through the secretory pathway. The signal recognition particle arrests the elongation of the polypeptide by binding the N-terminal signal sequence of the lysosomal protein. The ribosome in complex with the nascent polypeptide is then targeted to the ER membrane (Wiedmann *et al*., 1987; Ataide *et al*., 2011), where the Sec61-translocon co-translationally directs the nascent polypeptide into the ER lumen (Simon and Blobel, 1991; Osborne *et al*., 2005).

Folding of the nascent polypeptide begins as soon as the first part of it reaches the ER lumen (Kowarik *et al*., 2002) and the folding process is assisted by several molecular chaperones (Jansen *et al*., 2012). All lysosomal hydrolases have modifications, such as proteolytic processing, formation of disulphide bridges and glycosylation (Schröder *et al*., 2010). Even though many of the polypeptide

modifications happen co-translationally (Kowarik *et al*., 2002), they are commonly referred to as post translational modifications.

Cleavage of the signal sequence in the ER is the first proteolytic processing step for lysosomal proteins. Many lysosomal proteins have several additional proteolytic cleavages, which are required for formation of the mature protein (Schröder *et al*., 2010). These proteolytic cuts mostly appear in later parts of the transport route (Hansen *et al*., 2004; Moreland *et al*., 2005).

Disulphide bonds are covalent bonds between the side chains of two cysteine residues, which increase protein stability (Matsumura *et al*., 1989). Lysosomal hydrolases typically contain disulphide bonds, and the formation of protein-specific disulphide patterns is catalysed by the multifunctional protein disulphide isomerase. Protein disulphide isomerase is capable of both breaking and rearrangement of the disulphide bonds (Lambert and Freedman, 1985).

Eukaryotic proteins contain two major types of glycosylation, O- and N-glycosylation, which are named after their glycosidic linkages to the protein. In lysosomal proteins, O-glycosylation occurs mainly on membrane proteins, but the majority of soluble lysosomal proteins contain several N-glycans (Kollmann *et al*., 2005; Sleat *et al*., 2008). Protein N-glycosylation affects protein folding (Shental-Bechor and Levy, 2008; Hanson *et al*., 2009), solubility (Shental-Bechor and Levy, 2008), transport (Braulke and Bonifacino, 2009) and stability (Wormald and Dwek, 1999; Hanson *et al*., 2009).

N-glycosylation is a sequential process. It starts in the ER with the addition of a dolichylpyrophosphate activated oligosaccharide to the nascent polypeptide (Burda and Aebi, 1999). The preassembled oligosaccharide has a conserved structure consisting of three glucose (Glc), nine mannose (Man) and two *N*-acetylglucosamine (GlcNAc) residues (Figure 3) (Burda and Aebi, 1999). Membrane bound oligosaccharyl transferase transfers the activated oligosaccharide from the lipid carrier to the amide group of an asparagine residue (Roboti and High, 2012) (Figure 3). This asparagine needs to be part of the consensus sequence Asn-X-Ser/Thr, where X can be any amino acid except a proline (Figure 3) (Bause, 1983). Mouse glycoproteomic studies have shown that N-glycosylation on Asn-X-Cys and other motifs also occurs occasionally (Zielinska *et al*., 2010).

Most, but not all potential N-glycosylation sites are glycosylated *in vivo.* The activity of the oligosaccharyl transferase is affected by the local conformation of the glycosylation area (Bause, 1983; Zielinska *et al*., 2010). N-glycans are mostly

located in loops and turns on the protein surface but also in β-sheets (Zielinska *et al.*, 2010). Cysteine is enriched in the sequences surrounding the N-glycosylation sites, which are in use.

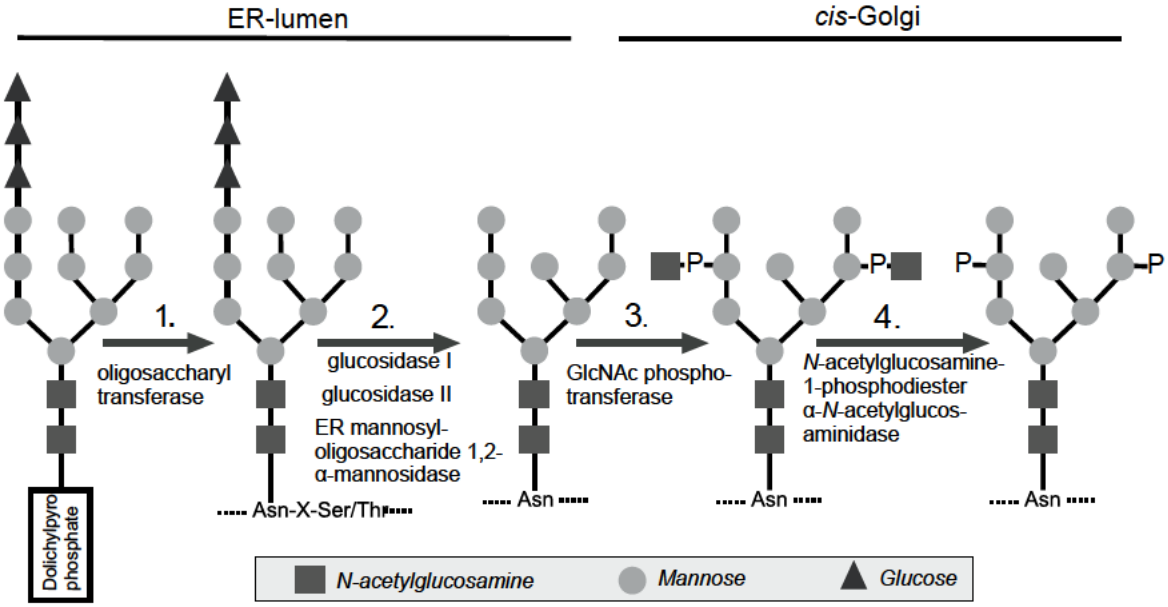


Figure 3. Schematic presentation of the main steps of the N-glycosylation and mannose 6 phosphorylation (M6P) of lysosomal hydrolases*. 1.The preassembled oligosaccharide on a dolichylpyrophosphate lipid carrier is transferred to the side chain nitrogen of asparagine in an Asn-X-Ser/Thr consensus sequence. 2.Before the protein leaves the ER lumen, the oligosaccharide is trimmed by sequential removal of the three glucose and the middle branch mannose residues. In the* cis *Golgi, specific glycans of lysosomal hydrolases are phosphorylated for the M6P-dependent pathway. 3. Phosphorylase adds a GlcNAc-phosphate to mannose residues. 4. The uncovering enzyme removes the GlcNAc residues. Diphosphorylation increases the affinity to M6P-receptor.*

2.2.2 The ER quality control

Most newly synthesised proteins leave the ER in cytosolic coatamer protein II coated vesicles, which bud off at the ER exit sites (Barlowe *et al.*, 1994; Venditti *et al.*, 2014). This process is strictly controlled and dependent on correct protein folding. Proteins can fail the ER quality control due to a truncation or mutation, which interferes with the proper protein fold. In addition, inability to form correct oligomers, lack of co-factor binding, missing or incorrect post translational modifications often lead to recognition by the ER quality control (Trombetta and Parodi, 2003). Misfolded proteins enter the ER-associated degradation (ERAD) system. From here, they are translocated into the cytosol, ubiquitinated and finally hydrolysed in the proteasome (Lederkremer, 2009).

N-glycosylation is tightly linked to glycoprotein folding and ER quality control. N-glycosylation happens co-translationally (Helenius and Aebi, 2004; Harada *et al*., 2009) and the N-glycosylation directly affects the folding process through the physico-chemical properties of the glycan (Shental-Bechor and Levy, 2008; Hanson *et al*., 2009).

The glycans also affect the protein folding process indirectly through the calnexin/calreticulin cycle (Hammond *et al*., 1994). The N-glycan modification starts during the translation when the outermost and the following glucose residues of $Glc_3Man_9GlcNAc_2$ are removed by glucosidase I and glucosidase II (Figure 3) (Aebi *et al*., 2010). The revealed $Glc_1Man_9GlcNAc_2$ is recognised by a membrane bound calnexin or its soluble homolog calreticulin (Caramelo and Parodi, 2008). These two ER resident lectins prevent the aggregation or premature ER exit of the polypeptide as it folds.

The exact composition of the protein glycan works as a signal that reflects the folding status of the protein (Helenius and Aebi, 2004). When glucosidase II removes the final glucose residue from the $Glc_1Man_9GlcNAc_2$ the protein is released from the lectin (Caramelo and Parodi, 2008). The enzyme UDP-Glc:glycoprotein glucosyltransferase Glucosidase II works as a conformation sensor and if the protein is incorrectly folded it reglucosylates the $Man_9GlcNAc_2$ glycan (Sousa and Parodi, 1995). As a result of the reglucosylation, the glycoprotein rebinds calnexin/calreticulin. This cycle of re- and deglucosylation associated with binding and release of calnexin/calreticulin continues until the protein is correctly folded or the folding is terminally unsuccessful.

When not bound to the calnexin/calreticulin, the N-glycans are exposed to the activity of the ER mannosyl-oligosaccharide 1,2-α-mannosidase. It removes the terminal mannose residue from the middle branch, resulting in $Man_8GlcNAc_2$ glycans (Figure 3) (Moremen and Molinari, 2006). Proteins staying longer in the re/deglucosylation cycle under further demannosylation as mannosyl-oligosaccharide 1,2-α-mannosidase continues removing mannose residues one by one resulting in $Man_{5-6}GlcNAc_2$ glycans, which are less attractive target for ER-exit and more likely to be passed to the ERAD system (Lederkremer and Glickman, 2005).

After the ER, the Man8GlcNAc2 glycans are further trimmed in the Golgi (Moremen *et al*., 1994; Moremen, 2002). These modifications are needed for transport of lysosomal proteins and they produce important additional structural complexity in the glycoproteins (Gabius *et al*., 2011).

2.2.3 M6P-dependent pathway

The mannose-6-phosphate (M6P)-dependent pathway is the best characterised and the major route for transport of soluble proteins to lysosomes. The M6P-tag is added to the glycans on lysosomal proteins in a two step process in the *cis*-Golgi (Pohlmann *et al*., 1982). The enzyme *N*-acetylglucosamine-1-phosphotransferase (GlcNAc phosphotransferase) recognises lysosomal proteins and transfers a GlcNAc-1-phosphate group to the C6 hydroxyl group of selected mannose residues on lysosomal proteins (Figure 3)(Bao *et al*., 1996). The number of phosphorylated N-glycans is protein specific and each N-glycan can be mono- or diphosphorylated (Varki and Kornfeld, 1980). A second enzyme, *N*-acetylglucosamine-1-phosphodiester α-*N*-acetylglucos-aminidase, removes the terminal GlcNAc residue thus uncovering the M6P group (Figure 3) (Kornfeld *et al*., 1999). The incompletely or non-phosphorylated polymannose glycans can be further modified to hybrid or complex type glycans (Moremen, 2002).

The labelled lysosomal proteins bind to the M6P-receptors in the *trans*-Golgi. The two M6P receptors, cation-dependent mannose-6-phosphate receptor (Olson *et al*., 1999) and cation-independent mannose-6-phosphate receptor (Zhou *et al*., 1995), complement each other, and are both required for lysosomal transport *via* the M6P pathway (Pohlmann *et al*., 1995). The M6P receptors and their cargo are transported through the endo-lysosomal pathway to lysosomes (Olson *et al*., 1999) (Figure 1). The pH of the endosomal vesicles is gradually decreased during the pathway, and at low pH of late endosomes, the M6P receptors release their cargo.

The M6P tags are rapidly removed soon after arrival of the protein in the lysosome (Makrypidi *et al*., 2012). The empty M6P receptors cycle back from the late endosomes to the *trans*-Golgi or to the cell membrane (Dahms *et al*., 2008). Mistakenly secreted M6P-tagged proteins are recaptured on the cell surface by the cation-independent mannose-6-phosphate receptors and with it they are then delivered through endocytosis back to the lysosome (Pohlmann *et al*., 1995). This feature is utilised in enzyme replacement therapy for lysosomal storage disorders (Brady, 2006).

Recognition of the lysosomal proteins for the M6P-dependent pathway is not based on a simple signal sequence. Instead, the recognition is dependent on the folded protein surface, which is recognised by GlcNAc phosphotransferase (Reitman and Kornfeld, 1981).

The recognition motif for GlcNAc phosphotransferase is known only for a limited number of lysosomal proteins (Table 1). In cathepsin D and L, two surface lysine

residues especially about 34 Å from each other, are important for the GlcNAc phosphotransferase binding (Cuozzo *et al.*, 1998; Warner *et al.*, 2002). In addition, there are other, lesser characterised recognition components (Steet *et al.*, 2005). Human N(4)-(β-*N*-acetylglucosaminyl)-L-asparaginase (AGA) also has surface lysines and at least one tyrosine residue that affect the level of mannose phosphorylation and AGA lysosomal transport (Tikkanen *et al.*, 1997). In arylsulfatase A, just one surface residue affected the phosphorylation level of the N-glycans, but similarly located lysine residues did not affect the N-glycan phosphorylation of the paralogous protein arylsulfatase B (Yaghootfam *et al.*, 2003).

Table 1. Experimentally verified GlcNAc phosphotransferase recognition signals.

Protein	**Phosphorylated N-glycans***	**Recognition***
Cathepsin D (Dittmer *et al.*, 1997; Cuozzo *et al.*, 1998; Steet *et al.*, 2005)	N134 (70) N263[&] (199)	K267 (203) K357 (293) K331 (267)
Cathepsin L (Cuozzo *et al.*, 1998; Warner *et al.*, 2002)	N221	K54[p], K99[p]; K157, K233, K237
N(4)-(β-*N*-acetylglucosaminyl)-L-asparaginase (Tikkanen *et al.*, 1995; Tikkanen *et al.*, 1997)	N38 N308	K177, Y178, K183[#] (154, 155, 160) K214 (191)
Arylsulfatase A (Yaghootfam *et al.*, 2003)	N158 N350	K457

*All sequences are numbered with UniProt sequence numbering, numbers in parenthesis are from the pdb-file. p= propeptide sequence, # area supported by a disulphide bridge. [&]Primary phosphorylation site.

2.2.4 M6P-independent pathway

Although the majority of the newly synthesised soluble lysosomal proteins follow the M6P-dependent pathway, there has long been evidence of alternate routes into lysosomes. The strongest evidence for M6P-independent lysosomal pathways comes from studies with cells from patients with Mucolipidosis II (known also as I-cell disease) (Coutinho *et al.*, 2012). Mucolipidosis II is caused by mutations in the gene encoding for the α/β-subunits of the GlcNAc phosphotransferase, with the result that all the lysosomal proteins lack the M6P tag (Tsuji *et al.*, 1988; Tiede *et al.*, 2005). The lysosomal proteins are thus unable to bind the M6P-receptors and should not reach the lysosomes at all. However, in certain cell types, some lysosomal hydrolases were still active, such as cathepsin D and glucosylceramidase (Rijnboutt *et al.*, 1991; Tsuji *et al.*, 1988).

To date two main receptors for the M6P-independent pathway have been identified, sortilin and lysosomal integral membrane protein type 2 (LIMP-2) (Coutinho *et al*., 2012). The mechanism of intracellular targeting of glucosylceramidase to lysosomes was unclear until Reczek and coworkers identified LIMP-2 association with glucosylceramidase (Reczek *et al*., 2007). As with other lysosomal membrane proteins, LIMP-2 is directed from the *trans*-Golgi to lysosomes by the cytosolic signal sequence and transported in clathrin-coated vesicles (Vega *et al*., 1991). LIMP-2 binds to soluble glucosylceramidase in a pH-dependent manner, and the release of glucosylceramidase occurs when LIMP-2 arrives in the acidic lysosome (Zachos *et al*., 2012).

Sortilin has been proposed to be involved in the transport of several lysosomal proteins such as prosaposin and Ganglioside GM2 activator, acid spingomyelinase and cathepsin D and H (Petersen *et al*., 1997; Lefrancois *et al*., 2003; Canuel *et al*., 2008). Sortilin is a multifunctional protein involved in many other processes in addition to the lysosomal transport, such as signal transduction (Nykjaer *et al*., 2004). It is highly expressed in many cell types, but especially in brain (Petersen *et al*., 1997).

2.3 Lysosomal proteins

The lysosomal proteome contains both soluble and membrane bound proteins, most of which are unique for lysosomes (Schröder *et al*., 2010). The hydrolytic function is concentrated in the lysosomal lumen and thus most lysosomal hydrolases are soluble proteins (discussed in more detail in section 2.3.2). In addition to the hydrolases, the lumen contains some other proteins that assist hydrolysis. The function of the lysosomal limiting membrane and the proteins within is to separate the lysosomal activities from the cell cytosol, control the fusion of lysosomes with other cell organelles and regulate the lysosomal content by bidirectional selective transport of molecules.

2.3.1 Proteins of lysosomal membrane

The lysosomal lipid bilayer has a specific phospholipid composition compared to other membranes in the animal cell (Gallegos *et al*., 2002) and its cholesterol concentration is lower than in the plasma membrane or late endosomes (Hamer *et al*., 2012).

The limiting lysosomal membrane contains at least 100 different proteins, including both transmembrane and membrane associated proteins (Schwake *et al*., 2013). Many of the lysosomal membrane proteins have been characterised only in recent years due to their hydrophobic nature and low abundance. The specific function of all of them is not yet clear, and many are multifunctional.

One of the key proteins in the lysosomal membrane is the vacuolar ATPase. It is a large multimeric proton-pump, which is essential for the lysosomal function as it keeps the interior of the lysosomes at pH of 4.6-5.0 (Mellman *et al*., 1986). The low pH makes the hydrolytic process more effective. The low pH optimum of lysosomal hydrolases also makes unwanted hydrolytic activity less likely during transport of nascent lysosomal proteins to lysosomes.

The most abundant lysosomal membrane proteins are the lysosome-associated membrane proteins-1 and -2 (LAMP-1 and LAMP-2). Because of their high abundance on lysosomal membranes in most tissue types, LAMP-1 and LAMP-2 are used as lysosomal markers in cell biology. LAMP-3 is a cell type specific protein found only in mature human dendritic cells (de Saint-Vis *et al*., 1998). LAMPs are heavily glycosylated and they form a glycocalyx-like coating on the lumenal side of the lysosomal membrane, which is believed to protect the lysosomal membrane from degradation (Wilke *et al*., 2012). LAMP proteins are also known to be involved in lysosomal biogenesis, phagocytosis (Eskelinen and Saftig, 2009) and in cholesterol transport (Schneede *et al*., 2011). Several other functions are probably still to be discovered.

LAMP-1 and LAMP-2 proteins are type I transmembrane proteins with one membrane spanning domain, a short cytosolic tail and a large lumenal domain. Glycosylation raises their molecular weight from the predicted 40-45 kDa to the observed 120 kDa (Carlsson *et al*., 1988). LAMP-2 pre-RNA can be alternatively spliced into three different isoforms that have changes both in transmembrane and cytosolic domains (Hatem *et al*., 1995). The structure of the membrane proximal domain of the human LAMP-3 is known (Wilke *et al*., 2012) and the LAMP-1 and LAMP-2 lumenal domains have been modelled based on the LAMP-3 structure(Wilke *et al*., 2012).

Lysosomal integral membrane protein type 2 (LIMP-2) is heavily glycosylated and the abundant type III transmembrane protein (Fujita *et al*., 1991) is involved in the delivery of glucosylceramidase from trans-Golgi to the lysosomes(Reczek *et al*., 2007).

Transporter proteins on the membrane enclosing the lysosome transport lysosomal substrates and degradation products across the membrane. For example, the limiting membrane contains several amino acid transporters, such as the proton-coupled amino acid transporter 1 (Sagné *et al*., 2001), which transports small neutral amino acids. Sialin is a transporter which was identified from Finnish disease heritage studies for Salla disease (Aula *et al*., 1979; Mancini *et al*., 1991). Sialin is a H^+ and sugar symporter, which exports sialic acid and acidic hexoses from lysosomes to the cytosol and imports asparagine and glutamine (Saftig and Klumperman, 2009).

The lysosomal membrane also contains a few enzymes, such as the heparan-α-glucosaminide *N*-acetyltransferase (Durand *et al*., 2010). This enzyme transfers an acetyl group from cytoplasmic acetyl-CoA to the terminal N-glucosamine residues of heparan sulphate within the lysosomes. The reaction is crucial for heparin degradation and lack of the activity causes lysosomal excess of stored heparin in mucopolysaccharidosis IIIC (Durand *et al*., 2010).

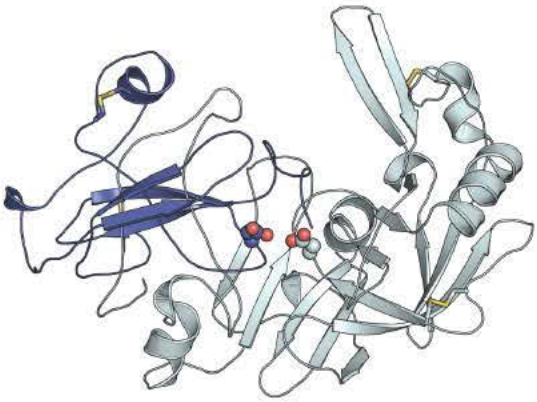
2.3.2 Soluble lysosomal hydrolases

The over 60 known soluble lysosomal proteins are more extensively characterised than the membrane embedded lysosomal proteins because they are often abundant and relatively easy to isolate. Several lysosomal proteins have been recognised through lysosomal storage disorders and genome studies. In addition, proteomic analysis has identified several novel candidates for soluble proteins in the lysosomal lumen (Kollmann *et al*., 2005; Schröder *et al*., 2010).

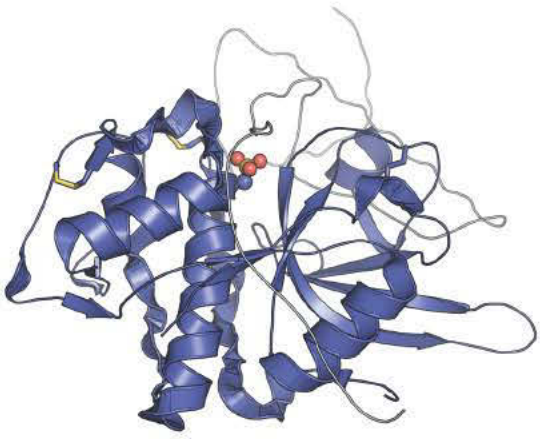
Lysosomal proteins degrade many types of macromolecules down to simple components such as monosaccharides, fatty acids and amino acids. Each hydrolase has a specific activity, and the degradation of most compounds is a highly organised multistep process, where the hydrolases work in a concerted manner. The lysosomal hydrolases can be classified on the basis of sequence or structural similarity, linked LSD (for example CLN1-8 proteins), catalytic activity or substrate (EC numbering). Based on substrate specificity, most lysosomal hydrolases fall into one the following groups: proteases, lipases, glycosidases, phosphatases, sulphatases and nucleases (Table 2).

Table 2. Soluble lysosomal hydrolases described in the section 2.3.2. Examples of monomer folds are in cartoon representation with selected catalytic residues shown in sphere. The peptides or domains if protein is formed from only one peptide are colored from the N-terminus to C-terminus with blue, turquise, violet, red and grey.

Substrate specificity	Example proteins	Involved in
Proteases		
	Tripeptidyl-peptidase 1 (UniProt O14773)	Degradation of proteins
Serine-cathepsin	Cathepsin A (UniProt P07339)	Degradation of proteins Transport of sialidase-1
Aspartic-cathepsin	Cathepsin D (UniProt P10619)	Degradation of proteins Cleavage of apoliprotein
Cysteine-cathepsin	Cathepsins B, C, F, H, K, S, L1 and L2 (UniProt P07858, P53634, PQ9UBX1, P09668, P43235, P25774, P07711, O60911)	Degradation of proteins Bone resorption (cat K)
	Below on the left the β-barrel fold of aspartic cathepsins and on the right the α-β complex fold of cysteine cathepsin. In cathepsin L1 the second peptide is the propeptide segment.	



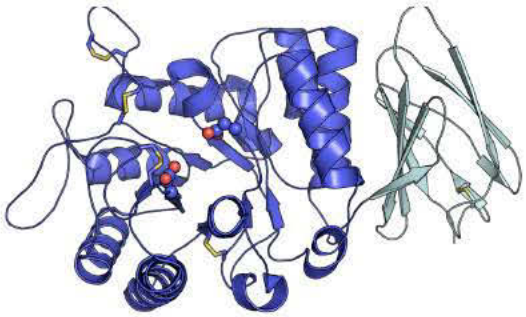
cathepsin D, pdb 1LYW,



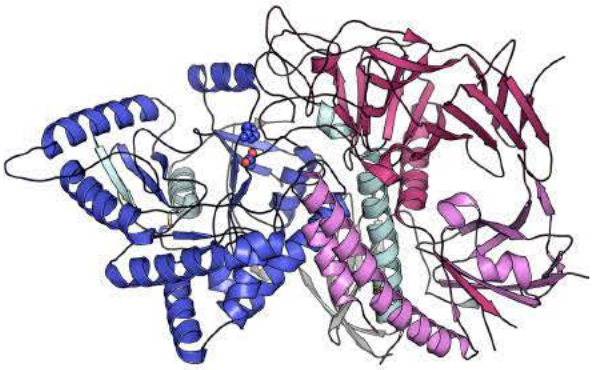
cathepsin L1, pdb 1CS8

Substrate specificity	Example proteins	Involved in
Lipases		
Phospholipases	Lysosomal Group XV phospholipase A (UniProt Q8NCC3)	Degradation of phospholipids
	Phospholipase D1 (UniProt Q13393)	Degradation of phospholipids
Sphingo-myelinases	Sphingomyelin phosphodiesterase (UniProt P17405)	Degradation of sphingomyelin
Ceramidases	Acid ceramidase (UniProt Q13510)	Degradation of ceramide
Cholesteryl esterase	Lysosomal acid lipase/cholesteryl ester hydrolase (UniProt P38571)	Degradation of cholesteryl esters
Glycosidases		
	Acid α-glucosidase (UniProt P10253)	Degradation of glycogen
	Hyaluronidase-1 (UniProt Q12794)	Degradation of hyaluronan
	α-L-iduronidase (UniProt P35475)	Degradation of dermatan and heparan sulphates
	β-glucuronidase (UniProt P08236)	Degradation of dermatan and keratan sulphates
	β-galactosidase (UniProt P16278)	Degradation of glycosphingolipids, glycoproteins, glycosaminoglycans
	Sialidase-1 (UniProt Q99519)	Degradation of oligosaccharides, gangliosides and glycoproteins
	β-hexoaminidases A, B, S (UniProt P06865 for subunit α, P07686 for subunit β)	Degradation of proteoglycans, glycolipids and glycoproteins
	Galactocerebrosidase (UniProt P54803)	Degradation of galactosphingolipids

Substrate specificity	Example proteins	Involved in
	Glucosylceramidase (UniProt P04062)	Glucosylceramidase (UniProt P04062)
	α-galactosidase A (UniProt P06280)	Degradation of glycolipids and glycoproteins
	Tissue α-fucosidase (UniProt P04066)	Degradation of glycoproteins
	Lysosomal α-mannosidase (UniProt O00754)	Degradation of glycoproteins
	β-mannosidase (UniProt O00462)	Degradation of glycoproteins
	Below on the left α-β barrel & sandwich fold of many glucosidases and on the right α-β barrel, up-down bundle, sandwich, distorted sandwich & sandwich fold of lysosomal α-mannosidase.	



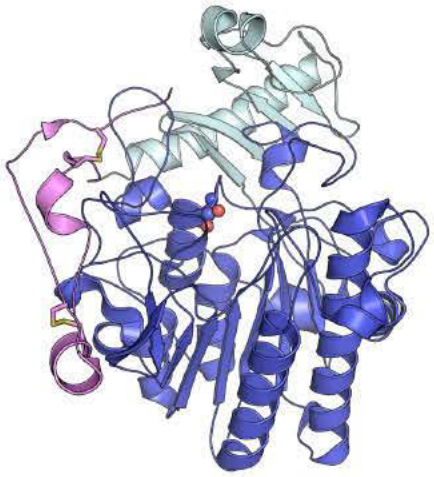
α-galactosidase A, pdb 1R46



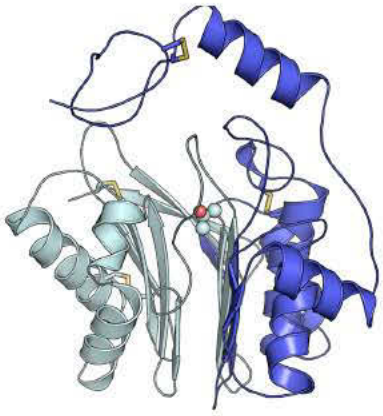
Lysosomal α-mannosidase, pdb 1O7D

Phosphatases		
	Lysosomal acid phosphatase (UniProt P13686)	Removal of the M6P-tag
	Tartrate resistant acid phosphatase type 5 (UniProt P13686)	Removal of the M6P-tag
Nucleases		
	Deoxyribonuclease-2-α (UniProt O00115)	Degradation of apoptotic nuclei

Substrate specificity	Example proteins	Involved in
Sulphatases		
	Arylsulfatase A (UniProt P15289)	Degradation of cerebrosides
	Arylsulfatase B (UniProt P15848)	Degradation of mucopolysaccharides
	N-acetylgalactosamine-6-sulphatase (UniProt P34059)	Degradation of mucopolysaccharides
Other		
Lipid carrier	Ganglioside GM2 activator (UniProt P17900)	Degradation of gangliosides
Aminidase	N(4)-(β-*N*-acetylglucosaminyl)-L-asparaginase (Uniprot P20933)	Degradation of glycoproteins
	α-N-acetylgalactosaminidase (UniProt P17050)	Degradation of glycoproteins
	Palmitoyl-protein thioesterase 1 (UniProt P50897)	Degradation of lipid modified proteins
	Phospholipase B-like proteins 1 and 2 (UniProt Q6P4A8 and Q8NH8)	Unknown
	Below on the left 3-layer (αβα) sandwich & 2-layer sandwich fold of the lysosomal sulphatases and on the right the 4-layer (αββα) sandwich fold of Ntn-hydrolases	



N-acetylgalactosamine-6-sulphatase, pdb



N(4)-(β-*N*-acetylglucosaminyl)-L-asparaginase, pdb 1APY

2.3.2.1 Proteases

Lysosomes contain more than 15 different proteases: they are involved both in the bulk degradation of proteins and in more specialised functions in many cellular processes (Müller *et al*., 2012). Most lysosomal proteases belong to the cathepsin family. Tripeptidyl-peptidase 1 (UniProt O14773) is one of the few lysosomal non-cathepsin proteases. It is a serine protease that cleaves tripeptides from the N-terminus of polypeptides in lysosomes (Lin *et al*., 2001).

Cathepsin proteases are divided based on sequence similarity into serine, aspartic and cysteine cathepsins and the same divisions hold also for structure and catalytic type (Rawlings *et al*., 2013). A few of the cathepsins, such as dipeptidyl-peptidase 1 (cathepsin C, UniProt P53634) are better known by their non-cathepsin name. Many lysosomal cathepsins are synthesised as procathepsins containing a propeptide (often a long loop) that needs to be removed in order for the enzyme to become catalytically active (Turk *et al*., 2012).

Cathepsin A (UniProt P10619) and cathepsin D (UniProt P07339) are the only known lysosomal members in their prospective cathepsin families. Cathepsin A, also known as lysosomal protective protein (Galjart *et al*., 1991) is a serine cathepsin. It is required for the transport of sialidase-1 to lysosomes and, together with β-galactosidase, the three proteins form a multi-enzyme complex (van der Spoel *et al*., 1998).

Cathepsin D is an aspartic cathepsin essential for instance in macroautophagy (Shacka *et al*., 2007) and is needed in the cleavage of apolipoprotein B in lysosomal low-density lipoprotein degradation (Van Lenten and Fogelman, 1990). It consists of two polypeptide chains linked by disulphide bridges (Baldwin *et al*., 1993). Cathepsin D has a typical aspartic protease fold (Table 2) with three domains; the central antiparallel β-sheet domain has N-and C-terminal lobes on both sides (Baldwin *et al*., 1993). Both of these lobes contain one N-glycan that can be phosphorylated (Cuozzo *et al*., 1998).

Cysteine cathepsins are members of the C1 family of papain-like enzymes (Turk *et al*., 2012) and the largest group of proteases in lysosomes (Rawlings *et al*., 2013). They include cathepsin B (UniProt P07858), cathepsin C (dipeptidyl-peptidase 1), cathepsin F (UniProt Q9UBX1), cathepsin H (UniProt P09668), cathepsin K (UniProt P43235), cathepsin S (UniProt P25774) and cathepsins L1 (UniProt P07711) and L2 (UniProt O60911). Cathepsin L2 is also known as cathepsin U or cathepsin V.

Most cysteine cathepsins are ubiquitously expressed, but the ones with more specialised roles can be expressed only in few tissues. For example, cathepsin B, which possesses both endopeptidase and dipeptidyl carboxypeptidase activities, is ubiquitously expressed (Rawlings *et al*., 2013), in contrast to cathepsin K, which is involved in bone resorption (Saftig *et al*., 1998). Cathepsin L1 and L2 have about 70 % sequence identity, but only L2 expression is restricted to thymus and testis (Coulombe *et al*., 1996; Brömme *et al*., 1999).

The papain fold of cysteine cathepsins is formed by two domains (Table 2), which are named left and right . The left domain contains three helices and the right domain is formed by a kind of a β-barrel (Turk *et al*., 2012). The active site catalytic histidine and cysteine residues and the residues interacting with the main chain of the substrate are conserved in the cysteine cathepsin family (Turk *et al*., 2012). The active site cleft is located at the interface of the two domains and it runs the full length of the interface in cysteine cathepsins that have predominantly endopeptidase activity (cathepsins F, L1, L2, and S)(Turk *et al*., 2001). In cathepsin C, the active site cleft is modified by an additional domain to provide exopeptidase specificity (Turk *et al*., 2001).

2.3.2.2 Glycosidases

Lysosomes receive carbohydrates as pure large polymers, such as glycogen and hyaluronan, or as glycoconjugates of lipids and proteins.

Some of the large polymeric carbohydrates, such as glycogen, are degraded by a single enzyme, in some occasions with a help of a few additional enzymes specialised for a particular glycosidic bond or substrate. The myriad of glycosidic linkages in the glycolipid and glycoprotein conjugates, requires several different glycosidases for a complete hydrolysis; only a handful of glycosidases act on both glycolipids and glycoproteins. Due to this, degradation of glycoconjugates follows a more strict sequential pathway than, for example, degradation of proteins. In addition, hydrolysis of the glycosidic bonds in glycolipids, requires several activator proteins, and the lipid moiety is usually left intact in lysosomes until all glycosidic bonds are hydrolased (Schulze and Sandhoff, 2011). Conversely, most lysosomal glycosidases have low activity on intact glycoproteins, suggesting that at least partial proteolysis must precede the action of glycosidases (Winchester, 2005).

Large polymer degradation

Glycogen degradation by lysosomal α-glucosidase is described in section 2.3.2.6.

Hyaluronan is a linear extracellular matrix glucosaminoglycan found in most tissues (Laurent and Fraser, 1992) and it is composed of disaccharide units of D-glucuronic acid and *N*-acetyl-D-glucosamine linked together with β(1->3) and β(1->4) glycosidic bonds. Of the five homologous human hyaluronidases, one is lysosomal. Hyaluronidase-1 (UniProt Q12794), cleaves the hyaluronan β(1->4) glycosidic bonds in lysosomes. Hyaluronidase-1 is endo-β-*N*-acetyl-hexosaminidase, and its activity results oligosaccharides of various lengths, which are probably further cleaved to monosachharides by β -exoglycosidases (Stern and Jedrzejas, 2006). Hyaluronidase-1 is formed by the N-terminal catalytic domain with a distorted $(\alpha/\beta)_8$ -barrel fold and a small C-terminal domain with an EGF like fold and three disulphide bridges.

α-L-iduronidase (UniProt P35475) and β-glucuronidase (UniProt P08236) are involved in mucopolysaccharide degradation. Mucopolysaccharides, also known as glucosaminoglycans are major components of the extracellular matrix. Mucopolysaccharides consist of repeating disaccharide units that form long unbranched chains. α-L-iduronidase hydrolases the nonreducing terminal α-L-iduronide glycosidic bond in the sulphated mucopolysaccharides dermatan sulphate and heparan sulphate (Scott *et al*., 1991).

Lipid conjugate deglycosylation

The lipid conjugates in eukaryotic cells are glycosphingolipids. They are composed of a carbohydrate headgroup linked by a β-glycosidic bond to a ceramide molecule. Cerebrosides are the simplest glycosphingolipids: they have only one sugar residue, which can be either glucose or galactose. Gangliosides have more complex carbohydrates, as their oligosaccharides are branched and contain at least three sugar residues of which one must be sialic acid. The sequential removal of sugar residues begins from the terminal residues.

β-galactosidase (Uniprot P16278) cleaves the terminal β-galactose of glycospingolipids such as GM1 ganglioside, glycoproteins and glucosaminoglycans (Distler and Jourdian, 1973). The β-galactosidase dimer has an N-terminal $(\alpha/\beta)_8$ domain and two smaller β-sheet domains (Table 2)(Ohto *et al*., 2012). β-galactosidase is part of the same multi-enzyme complex as cathepsin A (van der Spoel *et al*., 1998) and the complex is required for the proper processing and activity of β-galactosidase (Hoogeveen *et al*., 1986). The third member of the multi-enzyme complex, sialidase-1 (Uniprot Q99519), similarly requires formation of the complex for full activity. Sialidase-1 removes the terminal sialic acid

residues from oligosaccharides, gangliosides, and glycoproteins (Bonten *et al*., 1996).

The β-hexosaminidases A, B and S are dimeric enzymes that are formed of two subunits, α (UniProt P06865) and β (UniProt P07686), which have ~60% sequence identity to each other (Korneluk *et al*., 1986). Only the dimers are catalytically active (Maier *et al*., 2003). β-hexosaminidase A is an αβ heterodimer; β-hexosaminidase B is a β_2 homodimer; -and β-hexosaminidase S, an α_2 homodimer. β-hexosaminidases cleave terminal β-glycosidically linked *N*-acetylglucosamine or *N*-acetylgalactosamine residues from proteoglycans, glycolipids and glycoproteins (Maier *et al*., 2003). They are indispensable especially in the degradation of gangliosides, which also requies the ganglioside GM2 activator (UniProt P17900), a small lipid carrier (Wright *et al*., 2000).

α-galactosidase A (GLA, Uniprot P06280) hydrolases terminal α-galactose linkages in various glycoconjugates (Lemansky *et al*., 1987). The predominant glycolipid substrate of GLA is globotriaosylceramide, which has three galactose residues linked to each other by 1,4-glycosidic bonds. GLA is a homodimer with two domains in each monomer (Garman and Garboczi, 2004). The active site is a large opening located in the C-terminal end of the $(\alpha/\beta)_8$ barrel. The structure reveals that residues forming the active site make contacts to each functional group of the first galactose residue, but the protein has only little specificity for the binding of the residues following the glycosidic linkage (Garman and Garboczi, 2004). The second domain is a much smaller antiparallel β-domain formed by the C-terminus of the polypeptide (Garman and Garboczi, 2004).

Galactocerebrosidase (UniProt P54803) catalyses the final glycosidic step in catabolism of galactosphingolipids by hydrolysing the β-glycosidic bond between the galactose residue and the ceramide. One of the important substrates is β-D-galactocerebroside, the principal lipid component of myelin (Nagano *et al*., 1998). Similarly, glucosylceramidase (UniProt P04062) hydrolases glucosylceramide to ceramide and glucose (Sarmientos *et al*., 1986).

Glycoprotein deglycosylation

Glucosidic hydrolysis of the glycoprotein can start simultaneously both from the non-reducing and reducing ends. For complex and hybrid type N-glycans, this means removal of any of the fucose residues attached in the core GlcNAc or in the nonreducing end of the glycan by the enzyme tissue α-L-fucosidase (UniProt P04066) (Johnson and Alhadeff, 1991). The glycan residues are removed one by

one from the non-reducing end and many of the glycosidases catalysing the hydrolysis are the same as for other glycoconjugates, but some, such as lysosomal α-mannosidase (MAN2B1, Uniprot O00754) and β-mannosidase (UniProt O00462) only catalyse hydrolysis of glycoproteins.

MAN2B1 cleaves α-linked mannose residues from the non-reducing end of the N-linked glycoproteins (Tollersrud *et al*., 1997). The mature human MAN2B1 has four disulphide bridges and 11 N-glycans in its five proteolytic peptides (Tollersrud *et al*., 1997). Mature MAN2B1 is a dimer with five domains: an active site containing an N-terminal α/β-domain, an α-helical bundle and three β-domains (Table 2)(Heikinheimo *et al*., 2003). The MAN2B1 structure shows at least three interdomain ion bond networks, which could be involved in the low-pH activation of the protein (Heikinheimo *et al*., 2003). β-mannosidase cleaves the β-glycosidic bond between mannose and GlcNAc residues (Alkhayat *et al*., 1998).

The first cleavage in the reducing end is not hydrolysis of a glycosidic bond, but of an amide bond. The N-linked oligosaccharide is cleaved from the asparagine by N (4)-(β-*N*-acetylglucosaminyl)-L-asparaginase (AGA, Uniprot P20933) (Oinonen *et al*., 1995), if proteases have already freed the asparagine(Aronson, 1999). AGA has the N-terminal nucleophile (Ntn) amide hydrolase superfamily fold (Table 2) and a threonine residue as the N-terminal nucleophile (Oinonen and Rouvinen, 2000). It has two N-glycans, onto both of which M6P tag can be added (Tikkanen *et al*., 1995). α-*N*-acetylgalactosaminidase (UniProt P17050) catalyses the removal of the terminal α-*N*-acetylgalactosamine residue of the O-linked sugars attached to the serine or threonine residues of O-glycosylated proteins (Clark and Garman, 2009).

2.3.2.3 Lipases

Several phospholipases with different target bond activity or substrate specificity act in lysosomes. Group XV phospholipase A2 (Uniprot Q8NCC3) is a Ca^{2+}-independent lipase that hydrolyses phosphatidylcholine and phosphatidyl-ethanolamine (Hiraoka *et al*., 2002). Phospholipase D1 (UniProt Q13393) hydrolyses phosphatidylcholine to phosphatic acid and choline and has a role in exocytosis (Vitale *et al*., 2001).

Sphingomyelins contain a phosphocholine or phosphoethanolamine group attached to ceramide. Sphingomyelin phosphodiesterase (UniProt P17405) converts sphingomyelin to ceramide (Irvine *et al*., 1978). Acid ceramidase (UniProt Q13510) hydrolyses ceramide to sphingosine and free fatty acid (Gatt, 1963; Bernardo *et al*., 1995).

Cholesteryl esters are hydrolysed by lysosomal acid lipase/cholesteryl ester hydrolase (Uniprot P38571) (Du *et al*., 1998) and cholesterol is transported back to the cell membrane or to the ER (Möbius *et al*., 2003).

The full degradation of lipid containing molecules requires additional hydrolases such as palmitoyl-protein thioesterase 1 (UniProt P50897) (Bellizzi *et al*., 2000), which removes fatty acyl groups from cysteine residues in lipid-modified proteins.

2.3.2.4 Phosphatases, nucleases and sulphatases

Lysosomes contain at least two acid phosphatases: lysosomal acid phosphatase (Uniprot P11117) and the tartrate resistant acid phosphatase type 5 (UniProt P13686) (Suter *et al*., 2001). Lysosomal acid phosphatase is synthesised as a membrane bound precursor and proteolytically released in lysosomes. Both enzymes are monoesterases with a role in the removal of the M6P recognition marker from lysosomal proteins (Makrypidi *et al*., 2012).

The primary nuclease in lysosomes is deoxyribonuclease-2-α (UniProt O00115), which is proteolytically processed and activated in the lysosome (Ohkouchi *et al*., 2013). It is proposed to be the key enzyme in the degradation of apoptotic nuclei (Howell *et al*., 2003).

Compared to the phosphatases and nucleases, lysosomal sulphatases are relatively well characterised. They catalyse the hydrolysis of sulphate ester bonds. Four lysosomal sulphatases are known, of which three, arylsulfatase A (UniProt P15289), arylsulfatase B (UniProt P15848) and *N*-acetylgalactosamine-6-sulphatase (GALNS, UniProt P34059) are soluble. They are paralogous proteins with ~30 % sequence identity to each other.

The structures of all three are known and the overall folds are highly similar (Bond *et al*., 1997; von Bülow *et al*., 2001; Rivera-Colón *et al*., 2012). They are homodimers, where each monomer consists of three domains (Table 2)(Rivera-Colón *et al*., 2012). The active site domain has a core β-sheet, which is surrounded by α-helices. The active site is located in the center of the domain and has a modified cysteine residue that is required for activity. The second domain has an antiparallel β-sheet. The smallest domain is a"C-terminal meander", in which the last 41 residues form a loop structure stretching along the side of the active site domain and then threading half-way back towards the starting point of the loop.

Despite the structural similarity and the fact that these three enzymes have overlapping activities *in vitro*, they are non-redundant *in vivo* (Rivera-Colón *et al*., 2012). The main substrate of arylsulfatase A is a cerebroside 3-sulphate

(Lukatela *et al*., 1998). Arylsulfatase B is involved in the degradation of mucopolysaccharides such as dermatan sulphate. It hydrolyses the 4-sulphate groups in the *N*-acetylgalactosamine 4-sulphate residues of these molecules (Haskins *et al*., 1979). GALNS removes the 6-sulphate groups from the terminal *N*-acetylgalactos-amine-6-sulphate in mucopolysaccharides such as keratan sulphate (Rivera-Colón *et al*., 2012). The substrate specificity of the sulphates arises from the different overall sizes, shapes and electrostatistics of the active site pockets (Rivera-Colón *et al*., 2012).

2.3.2.5 Phospholipase B-like proteins

In mammals, the *PLBD1* and *PLBD2* genes encode paralogous soluble proteins with unknown cellular functions. These proteins, phospholipase B-like protein 1 (PLBD1, UniProt Q6P4A8 for the human protein) and phospholipase B-like protein 2 (PLBD2, UniProt Q8NHP8), are conserved and have homologs in several higher eukaryotes. They do not exist in plants, fungi or prokaryota. The sequence identity between human PLBD1 and PLBD2 is approximately 30 %. They also both have ~30 % sequence identity to a *Dictyostelium discoideum* phospholipase B-like protein (UniProt Q550U9).

The phospholipase B-like protein family is named after the amoeba protein which was shown to have low phospholipase B (PLB) activity (Morgan *et al*., 2004). PLB proteins hydrolyse fatty acids both from the *sn1* and *sn2* positions in glycerophospholipids (Figure 4) (Lee *et al*., 1994). So far, the *Dictyostelium discoideum* protein has remained the only phospholipase B-like protein that has been shown to have PLB or indeed any other type of enzymatic activity (Jensen *et al*., 2007; Deuschl *et al*., 2006).

PLBD1 is a soluble glycoprotein that was first purified from human secretory lysosomes (Xu *et al*., 2009). It has also been identified in several proteomic studies of lysosomal proteins (Sleat *et al*., 2005; Sleat *et al*., 2008; Della Valle *et al*., 2011; Chen *et al*., 2009). Furthermore, the human PLBD1 contains M6P-residues (Sleat *et al*., 2008), which is why it was considered a candidate lysosomal protein (Schröder *et al*., 2010).

The human neutrophile derived PLBD1 formed two fragments on SDS-PAGE with molecular weights of 22 kDa and 42 kDa. After preincubation for several weeks, the protein showed PLB activity, but significantly lower than that reported for other PLBs. Xu *et al* (2009) proposed that preincubation is necessary for the activation of the 42 kDa fragment and that activation requires a proteolytic cut in N- and C-terminus of the 42 kDa fragment.

PLBD2 (also called 66.3 kDa protein) was identified in two independent lysosomal proteomic studies (Journet *et al.*, 2002; Kollmann *et al.*, 2005). Both the human and mouse PLBD2 (hPLBD2, mPLBD2) were characterised following their identification (Jensen *et al.*, 2007; Deuschl *et al.*, 2006). Transiently transfected human and mouse PLBD2 (Jensen *et al.*, 2007) and endogenous mPLBD2 (Deuschl *et al.*, 2006) localise in lysosomes under immunofluorescence studies. In addition, the mouse liver subcellular fractionation showed mPLBD2 in the lysosomal fraction (Jensen *et al.*, 2007; Deuschl *et al.*, 2006).

Similarly to PLBD1, hPLBD2 and mPLBD2 are formed of two main fragments with relative molecular masses of approximately 32 kDa and 50 kDa for hPLBD2 (Jensen *et al.*, 2007) and 28 kDa and 40 kDa for mPLBD2 (Deuschl *et al.*, 2006).

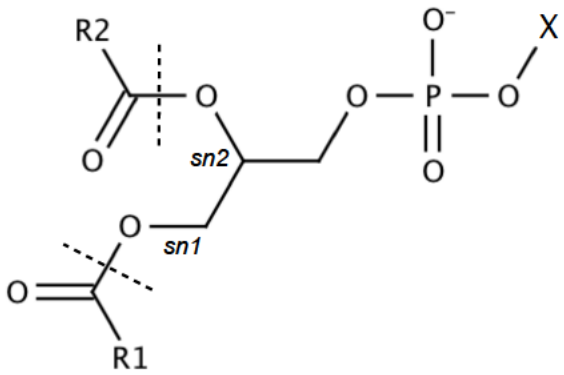


Figure 4. Molecular formula of glycerophospholipds*. Glycerophospholipids consist of glycerol-3-phosphate esterified to two fatty acid chains (R1 and R2). The phosphoryl group is esterified to a polar group X, which can be hydrogen, ethalolamine, choline, serine, myo-inisitol, glycerol or phosphatidylglycerol. The fatty acids attached to the sn1 position are typically saturated 16 or 18 carbon long fatty acids and in sn2, unsaturated fatty acids containing from 16 to 20 carbons. The bonds that PLBs are able to cleave are marked by dashed lines.*

All putative N-glycosylation sites, five in mPLBD2 and six in hPLBD2, were occupied (Deuschl *et al.*, 2006; Jensen *et al.*, 2007).

The three dimensional structure of mPLBD2 is known (PDB 3FBX; Lakomek *et al.*, 2009). The fold of the mPLBD2 structure belongs to the N-terminal nucleophile (Ntn) aminohydrolases (Ntn-hydrolase) superfamily (Lakomek *et al.*, 2009). Ntn-hydrolases share a similar four-layered αββα-fold and conserved catalytic machinery (Oinonen and Rouvinen, 2000), but no recognisable sequence identity. They all catalyse amide bond hydrolysis, on a variety of substrates from small antibiotics to full proteins. Ntn-hydrolases use only one amino acid in the

catalysis (Oinonen and Rouvinen, 2000). This catalytic N-terminal nucleophile is created autoproteolytically from the precursor protein. The N-terminal nucleophile can be either a threonine, serine or cysteine residue and its α-amino group works as a catalytic base in the reaction. The preserved stereochemistry of the active site suggests that all Ntn-hydrolases share the same catalytic mechanism (Oinonen and Rouvinen, 2000).

The proposed lipase function for the PLBD family is inconsistent with the Ntn-fold, since all the other Ntn-family enzymes cleave amide not ester bonds (Oinonen and Rouvinen, 2000).

2.3.2.6 Lysosomal α-glucosidase

Lysosomal α-glucosidase (GAA Uniprot P10253, also known as acid α-glucosidase) is the sole enzyme in lysosomes responsible for degrading glycogen to glucose (Kroos *et al*., 2012). GAA is unrelated to the enzymes of the cytosolic glycogen degradation (see below) and it hydrolyses both α(1->4) and α(1->6) glycosidic bonds, unlike the cytosolic glycogen phosphorylase.

Glycogen molecules are large multi-branched polymers of glucose (Figure 5). Glycogen is an efficient, quickly-mobilised form of energy stored in animal cells (Roach *et al*., 2012). Glycogen is principally made and stored in liver and muscle cells, where it is the main energy source in high intensity exercise (Romijn *et al*., 1993). Enzymatic dephosphorylation of glucose is required for glucose release through the cell membrane (Roach *et al*., 2012). Thus only liver cells can participate in regulation of blood glucose levels, as only they can release D-glucose back into the blood stream.

Cellular glycogen is stored in the cytosol as granules, which consist of a single glycogen molecule and complement proteins. The complement proteins are responsible for the autonomous synthesis and breakdown of each glycogen molecule (Shearer and Graham, 2004). In addition to the classic cytosolic breakdown, where glycogen is cleaved through glucose-phosphate derivatives, glycogen is converted directly to glucose in lysosomes. Lysosomal degradation requires that some of the cytosolic glycogen is transported to lysosomes. The exact mechanism of the transport is still unclear (Raben *et al*., 2008), but it is proposed to resemble selective autophagocytosis (Jiang *et al*., 2010).

Lysosomal degradation of glycogen is crucial for cells. The glycogen storage disease type II, Pompe disease, is caused by mutations in the *GAA* gene. Insufficient GAA activity causes progressive accumulation of glycogen in

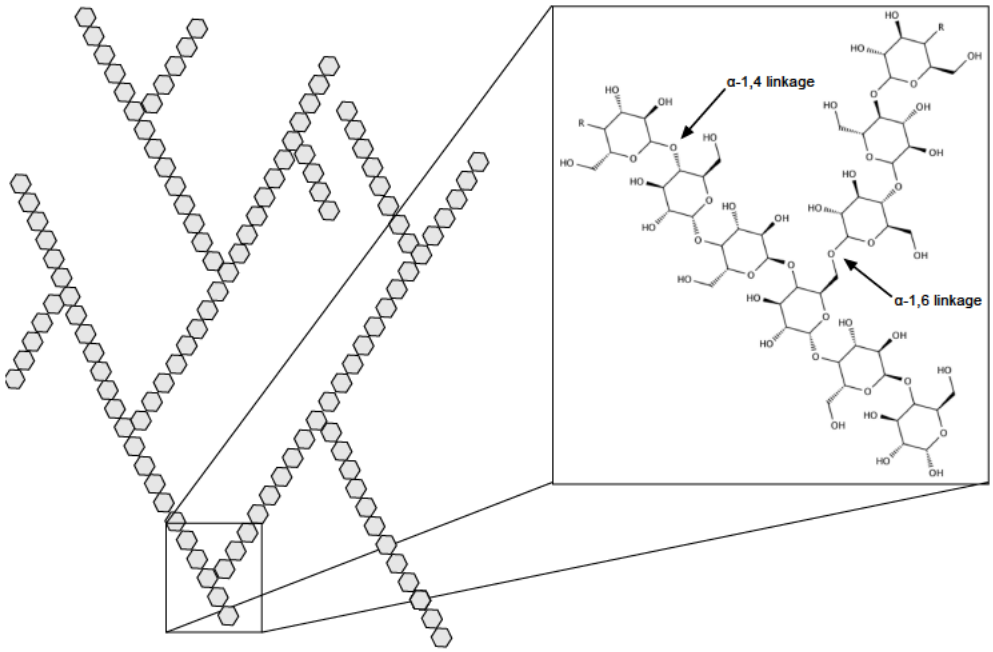


Figure 5. Structure of glycogen. *The schematic drawing on the left side shows a section of a branched glycogen polymer with the glucose residues as grey hexagons. Inset on the right demonstrates the two type of glycosidic bonds; α(1->4) glycosidic bonds connect the glucose residues in the linear polymer and α(1->6) glycosidic bonds connect a branch to the linear polymer. In average, a chain is 13 residues long and has two branches (Meléndez et al., 1997). Branching increases glycogen solubility and the rates of both synthesis and degradation.*

lysosomes (Beratis *et al.*, 1978). The *GAA* gene (GenBank NM_000152) located on chromosome 17 consists of 20 exons, 19 of which encode the 952 amino acid long GAA precursor. The molecular weight of the precursor is 110 kDa and it includes seven N-glycans (Moreland *et al.*, 2005). The N-glycan moieties are modified and the peptide precursor proteolytically processed in a stepwise manner *en route* to the lysosomes, resulting in two separate lysosomal mature 76/70 kDa GAA forms, which are difficult to experimentally separate from each other (Moreland *et al.*, 2005).

The 76/70 kDa forms are named after the apparent molecular mass of the largest peptide found in the mature protein. The 76 kDa form consists of three peptides (Moreland *et al.*, 2005). The smallest, 3.9 kDa peptide (amino acids 78-113) has no N-glycans and is attached with a disulphide bridge to the 76 kDa peptide (Moreland *et al.*, 2005). The 76 kDa peptide (residues 122-782) has five N-glycans and the third, 19.4 kDa peptide (residues 792-952) has the final two N-glycans. The 70 kDa form consists of four peptides as a result of an additional proteolytic cleavage in the 76 kDa peptide, which results in a 10.4 kDa peptide (residues

122-199) and a 70 kDa peptide (residues 204-782). The 3.9 kDa peptide is thus in the mature 70 kDa form attached by a disulphide bridge to the 10.4 kDa peptide instead of the largest peptide (Moreland *et al*., 2005).

The 76/70 kDa forms have been reported to have a 7-10 increase in glycogen affinity compared to the 110 kDa precursor form, and it has been thus suggested that proteolytic processing is needed for optimal enzymatic activity (Wisselaar *et al*., 1993; Bijvoet *et al*., 1998).

GAA is an exohydrolase and by sequence similarity belongs to the glycoside hydrolase family 31 (GH31). GH31 enzymes are a heterogeneous group of glycoside hydrolases that have broad activities and a wide range of substrate specificities (Henrissat and Davies, 1997). For instance, the human sucrase/ isomaltase, human intestinal maltase-glucoamylase and α-galactosidase (GLA) are all GH31 enzymes. The N-terminal domain of the intestinal maltase-glucoamylase (UniProt O43451) has 44% sequence identity with GAA and its three dimensional structure has been solved (PDB 2QLY) (Sim *et al*., 2008). The catalytic residues of GH31 enzymes include two aspartic acids, one functioning as a nucleophile and the other as a general acid/base. The region surrounding the aspartic acid residues is conserved in GH31 enzymes (Sim *et al*., 2008).

The catalytic base in human GAA is Asp518 (Hermans *et al*., 1991). Asp518 is located in the 76/70 kDa peptide of the mature GAA. The structure of GAA is unknown, but a GAA homology model based on the intestinal maltase-glucoamylase structure suggests that GAA is composed of three domains (Bruckmann *et al*., 2012). In the model, the active site of GAA is formed on the top of the central $(\alpha/\beta)_8$ barrel (Bruckmann *et al*., 2012).

2.4 Significance of lysosomes in human health

Due to the indispensable role of degradation and recycling for cell homeostasis and for the various more specific lysosomal roles, proper function of the lysosomal system is vital for living cells. Failure of lysosomal function has been long known to result in lysosomal storage disorders (LSDs). In recent years the importance of the intact lysosomal pathways has become evident in diseases other than the orphan LSD's.

LSD mutation carriers appear to be more susceptible to common diseases, such as stuttering, dementia, Parkinson's and Alzheimer's diseases (Kang *et al*., 2010; Boya, 2012; Nalls *et al*., 2013; Avrahami *et al*., 2013). The best studied example of

this is the connection between Gaucher disease and Parkinson's disease. The recessively inherited Gaucher disease is caused by mutations in the *GBA* gene encoding a soluble lysosomal hydrolase, Glucosylceramidase. Recessive mutations in *GBA* predispose to Parkinson's disease and it is in fact the most common genetic risk factor for Parkinson's disease (Tayebi *et al*., 2003; Sidransky *et al*., 2009; Nichols *et al*., 2009; Dehay *et al*., 2013).

The exposure of the link between the LSD mutations and common diseases has significance on several levels. The studies have already given insight to the function of the lysosomal system and to the mechanism of the disease pathologies of both the LSDs and Parkinson's disease (Dehay *et al.*, 2013; Schapira and Gegg, 2013).

Lysosomal dysfunction has an important role in the mechanism of Parkinson's disease pathogenesis (Osellame *et al*., 2013). For instance, the Gaucher disease mouse model has shown that dysfunctional mitochondria and insoluble α-synuclein accumulate to the cytosol of the cell as a result of impaired autophagic and proteosomal machineries (Foltynie and Kahan, 2013). These findings open up possibilities for new therapeutic approaches to Parkinson's disease and related synucleinopathies, such as augmenting glucosylceramidase activity (Sardi *et al*., 2013). Also carriers of the *GBA* risk mutations could benefit from treatments before any Parkinson's disease symptoms appear (Schapira and Gegg, 2013).

Lysososomes have an important role in the progression of common diseases that do not involve identified LSD mutations. For instance, recent studies indicate that lysosomal dysfunction caused by deficient lysosome mediated autophagy could play a part in the progression of diabetic neuropathy (Peres *et al*., 2013). Thus, autophagy activation is a possible therapeutic target for diabetic neuropathy (Kume *et al*., 2014).

The role of lysosomes in cancer progression is dual. Some of the cancer-associated lysosomal changes contribute to tumor growth and some are tumor suppressing (Kallunki *et al*., 2013). The expression of cathepsins *in vitro* is up-regulated in cancer cells and thus the inhibition of cysteine cathepsins has potential in cancer therapy (Kallunki *et al*., 2013). For instance, cathepsin B is a key contributor to bone metastasis and its inhibition significantly reduces metastasis (Withana *et al*., 2012).

2.4.1 Lysosomal storage disorders

The over 50 known LSDs are inherited diseases that often lead to severe disabilities and premature death. They are caused by a mutation in a single gene encoding a lysosomal or lysosome related protein (Hopwood, 2012). Common to all LSDs is impaired lysosomal degradation and the subsequent progressive storage of unhydrolysed material or digestion products in lysosomes. This results in damage to multiple cellular systems and tissues (Platt *et al*., 2012; Cox and Cachon-Gonzalez, 2012). LSDs are often grouped according to the chemical nature of the primary storage material and these groups include for instance lipidoses, mucopolysaccharidoses and neuronal ceroid lipofuscinoses.

The individual LSDs are rare diseases, but collectively they form significant group of inherited diseases, which affect 1 out of every 7700 newborn children (Meikle *et al.*, 1999). The mode of inheritance in most LSDs is autosomal recessive, but some such as Fabry disease have X-linked inheritance.

In a classical LSD, the disease is caused by mutations in a gene encoding a lysosomal hydrolase and the material building up to lysosomes is the substrate of this hydrolase. Pompe disease, α-mannosidosis, Morquio A syndrome, aspartylglucosaminuria, Gaucher's disease, Fabre's disease and sialidosis are examples of classical LSDs. Reduced or missing protein activity in lysosomes can occur in several ways, related to the type of the disease-associated mutation. The mutation can prevent the synthesis the hydrolase, lead to truncated protein, prevent correct folding of the enzyme or lead to a protein with disturbed active site (Kuokkanen *et al*., 2011; Rivera-Colón *et al*., 2012). In addition, the protein may be insufficiently transported or unstable in lysosomes (Kuokkanen *et al*., 2011; Rivera-Colón *et al*., 2012).

However, as LSDs form a diverse group of diseases (Hopwood, 2012), even the 'classic' LSDs cannot be considered as 'typical' LSDs. In several LSDs, the gene containing an LSD-associated mutation, encodes for a protein, which is not a lysosomal hydrolase, but a protein related to the formation or function of lysosomes (Sleat *et al*., 2009). Some LSDs have still unknown etiology (Sleat *et al*., 2009). Thus the compounds accumulating in the lysosomes are not always substrates of a single lysosomal hydrolase. In Salla disease the protein with impaired function is a transporter (Aula *et al*., 1979). Sialic acid, which builds up in lysosomes in Salla disease, is an end product of lysosomal hydrolysis, which is not transported further (Mancini *et al*., 1991).

Mucolipidoses II and III are caused by impairement in the lysosomal targeting machinery. GlcNAc phosphotransferase is required for addition of M6P-tag and thus for transport of most lumenal lysosomal proteins to lysosomes. It is a heterohexameric complex formed of three subunits; α, β and γ. Mucolipidosis II is caused by pathogenic mutations in the gene *GNPTAB* encoding GlcNAc phosphotransferase α and β subunits (Tiede *et al*., 2005). Because of the impaired GlcNAc phosphotransferase function, the lysosomal activity of most lysosomal hydrolases is affected in Mucolipidoses II (Hickman and Neufeld, 1972). Mutations of the *GNPTG* gene encoding the soluble γ-subunit are associated with the clinically less severe mucolipidosis III (Raas-Rothschild *et al*., 2000).

Gradual accumulation of lysosomal storage material in LSDs disturbs the function of the whole cell by a complex mechanism, which involves a cascade of numerous secondary cellular effects triggered by the storage of the material (Cox and Cachon-Gonzalez, 2012). In most LSDs, the lysosomal storage of the material occurs in all cell types, but depending on the disease, some organs and tissues are more affected than others (Boustany, 2013). The clinical manifestations of LSDs differ widely between different disorders, but often heart bone, muscle, liver, kidney and spleen are affected. Approximately 2/3 of all LSD patients have central nervous system involvement and progressive neurodegeneration (Micsenyi and Walkley, 2012).

Due to the heterogeneity of the disease-associated mutations and the multi-organ involvement in many LSDs, even patients with the same LSD have variation in the clinical symptoms. Most LSDs include an early, more severe and a late, milder form of the disease. In the early form, accumulation of compounds in the lysosome starts at birth or in early childhood. The early onset phenotypes often lead to death at an early age. In the late onset phenotypes, the lysosomal storage material appears more slowly and the symptoms can appear during childhood or during adulthood.

LSDs are characterised by low enzyme activity in the lysosomes of the patients cells. Each tissue type has a disease specific threshold for enzyme activity and under this threshold the effects of the mutation are pathogenic (Marsden and Levy, 2010). Enzyme activity measurement of patient isolated cultured fibroblasts, is often used in diagnosis and in prediction of the prognosis of the disease (Winchester, 2012). In general, early onset patients have no or very little enzyme activity, although low activity may also be connected with late onset patients (Willemsen *et al*., 1993). Depending on inherited mutation and patient phenotype,

even a moderate therapeutic increase in the lysosomal enzyme activity can significantly improve the quality of life for the patient.

In genotype-phenotype correlation, mutations which cause large disturbance to the protein structure or mutations which directly affect the active site residues, are usually early onset causing phenotypes. However, direct linking of the genotype to the phenotype, may be challenging (Kroos *et al*., 2012). Especially, distinction of late onset disease-associated mutations from polymorphism is difficult (Desnuelle and Salviati, 2011). This is an important issue in progressive diseases, as many late onset patients would benefit from early treatment, before the first symptoms emerge (Toscano *et al*., 2013). Disease prevention has great potential in the future, as more LSD mutations are screened on population level from newborn babies (Labrousse *et al*., 2010; Marsden and Levy, 2010; Yang *et al*., 2014) or when more carriers of the LSD mutations are linked with common diseases of the elderly (Nichols *et al*., 2009; Dehay *et al*., 2013).

2.4.1.1 Pompe disease

Pompe disease, also known as glycogen storage disease type II, was the first LSD described (Hers, 1963). Its mode of inheritance is autosomal recessive. Mutations in the *GAA* gene encoding lysosomal α-glucosidase cause impairment of the GAA activity in the lysosomes. GAA activity of Pompe patients can be either severely lowered or completely absent (Beratis *et al*., 1978; Kroos *et al*., 2008). Insufficient GAA activity leads to lysosomal accumulation of the GAA substrate, glycogen (Hers, 1963; van der Ploeg and Reuser, 2008).

Like in other LSDs, the birth prevalence of Pompe disease varies by population. Predictions of the birth prevalence are between 1 in 40 000 newborns in the Netherlands (Ausems *et al*., 1999) and 1 in 600 000 newborns in Northern Portugal (Pinto *et al*., 2004). In Finland, only three Pompe disease cases have been diagnosed (Vanto *et al*., 1982; Korpela *et al*., 2009) suggesting that the birth prevalence of Pompe in Finland is much lower than the overall estimated prevalence of Pompe disease in Western countries (Korpela *et al*., 2009).

The symptoms of Pompe disease are caused mainly by damage of cardiac, skeletal and smooth muscle (van den Hout *et al*., 2003). Unlike LSDs in general, Pompe disease less frequently affects the central nervous system (Boustany, 2013). An important contributor to the muscle weakness is the disordered intracellular recycling system. This does not involve only the lysosomal degradation, but also UPS and phagocytosis are affected (Raben *et al*., 2008).

Early age onset Pompe patients have symptoms before the age of one and there is cardiac involvement, which is absent in late onset patients (Kroos *et al*., 2008). In the late onset Pompe phenotypes, the symptoms mainly include progressive muscular weakness, which appears during childhood or during adulthood (Kroos *et al*., 2012).

2.4.2 Current therapies for lysosomal storage disorders

Most treatments for LSDs still concentrate on managing the disease-associated symptoms, but several disease-specific treatments have emerged over the past two decades (Parenti *et al*., 2013).

One of the first disease-specific treatments, which is still in use, was hemopoietic stem cell transplantation (Giugliani *et al*., 2010). In 1980, the first bone marrow transplant was done to a Mucopolysaccharidose I patient to endogenously supply the deficient α-L-iduronidase with a normal one (Hobbs *et al*., 1981). The treatment depends on finding suitable healthy bone marrow donors (Valayannopoulos and Wijburg, 2011).

In gene therapy, as in hemopoietic stem cell transplantation, the aim is to restore the lysosomal enzyme activity endogenously by introducing a normal copy of the defected gene into a subset of the patient's cells (Parenti *et al*., 2013). Several methods, such as transfecting the patient's cells by viral vectors (Seregin and Amalfitano, 2011), or replacing subset of a specific cell type with genetically corrected stem cells or microencapsulation technique (Boelens and Wynn, 2013), have been tested for the gene transfer. In all these cases, the transfected cells will provide a systemic pool of the correct enzymes (Boelens and Wynn, 2013). The development of gene therapy has been successful in many animal tests, but also encountered problems, such as the immunogeneity produced by the viral vectors (Seregin and Amalfitano, 2011). Thus the great potential of gene therapy has not yet been successfully turned into an approved treatment in any of the LSDs (Parenti *et al*., 2013).

In enzyme replacement therapy (ERT), the deficient enzyme is replaced exogenously. The intravenously introduced recombinant enzyme is targeted to lysosomes through the M6P-receptors on the cell surface (Brady, 2006). ERT is an approved treatment for six LSDs and is in developmental stage for several more (Ohashi, 2012). In Gaucher disease and in Fabry disease, ERT became available in 1991 and in 2001 respectively and for Mucopolysaccharidosis types I, II and VI in 2003, 2007 and 2006 (Glamuzina *et al*., 2011; Wraith *et al*., 2005; Lourenço and

Giugliani, 2014). For Pompe disease, ERT with recombinant human GAA (rhGAA, Myozyme) was approved in 2006. It is still the sole treatment for Pompe disease (van den Hout *et al*., 2001; Kishnani *et al*., 2007).

ERT prolongs life and reduces symptoms of the Pompe patients especially among early onset patients when treatment is started early (Kishnani *et al*., 2007). In adult patients as well, the treatment improves the skeletal function to some extent and prevents respiratory failure (van der Ploeg *et al*., 2010; Orlikowski *et al*., 2011), even though clearance of the stored glycogen is not as effective as in early onset patients. This is partly thought to be because the impaired phagocytosis caused by the prolonged storage of glycogen (Raben *et al*., 2008).

The high price and other drawbacks of ERT has led to interest in alternative treatments. In substrate reduction therapy, the biosynthetic pathway of the substrate of the defective enzyme is inhibited with small molecules to reduce the amount of the substrate (Valayannopoulos and Wijburg, 2011). Substrate reduction therapy is especially valuable in diseases which have strong central neurosystem involvement, as the blood-brain barrier creates difficulties for drug delivery, especially of the large recombinant proteins used in ERT (Boustany, 2013).

Enzyme enhancement therapy (EET) is another treatment involving small molecules (Porto *et al*., 2009). Most LSD associated mutations cause destabilisation of the target protein fold and lead to the ERAD degradation of the newly synthesised enzyme. In EET, competitive inhibitors work as chemical chaperones stabilising the protein molecules containing destabilising mutations and thus prevent their premature degradation (Fan, 2008). EET is in clinical trials for Fabry disease and Gaucher disease (Shayman and Larsen, 2014).

EET has shown great potential for Pompe disease. Several competitive inhibitors of GAA could function as a base for drug development for Pompe (Porto *et al*., 2009; Yoshimizu *et al*., 2008; Flanagan *et al*., 2009; Parenti *et al*., 2007). In addition, an allosteric, not a competitive inhibitor, could be for the first time used as a chemical chaperone in EET on GAA (Porto *et al*., 2012). EET has several advantages in the treatment of Pompe disease in comparison to ERT. It is much cheaper, admitted orally and reaches all organs more equally than rhGAA. In cases where the disease-associated mutation is located in the active site and EET does not thus increase the residual lysosomal GAA activity, it can be used in combination with ERT to stabilise the recombinant enzyme and decrease the amount of needed rhGAA (Porto *et al*., 2009; Porto *et al*., 2012).

3 Aims of the study

For the present study, I worked with two soluble lysosomal proteins and with an analysis of known structures of soluble lysosomal proteins. When the work was initiated, phospholipase B-like protein 1 (PLBD1) was not well characterised: its lysosomal targeting was only hypothesised and its function was unknown. The second target protein in this study, acid α-glucosidase (GAA), has been previously extensively studied, not least because it is mutated in Pompe disease. Despite this, no three dimensional structure was known for GAA or its close relatives.

The aims of the study were:

-To crystallise and solve the structure of bovine PLBD1 and human GAA.

-To propose a function for the PLBD1 protein and the lysosomal PLBD family.

-To understand the relationship between rhGAA and its competitive inhibitors to propose suitable chaperone lead compounds for GAA.

-To better understand the structural basis of lysosomal protein structure and their targeting to lysosomes.

-To analyse the molecular mechanism of lysosomal storage disorder associated mutations.

4 Methods

4.1 Methods used in studies I-III

Methods used in this thesis are listed in Table 3. Detailed descriptions can be found in the original publications I-III. The method for protein crystallisation, which is unpublished work, is described in section 4.2.

Table 3. Methods used in the original publications.

Method		Study
Protein analysis	Protein purification	II
	SDS-PAGE	I, II
	Protein concentration measurements	I, II
	N-terminal sequencing	II
	Deglucosylation	II
	Enzyme activity assay	I
Thermal stability measurements	Differential scanning fluorimetry	I, II
Subcellular localisation	Immunofluorescence microscopy	II
Bioinformatics	Sequence comparisons	II, III
	Surface exposure evaluation for residues	III
	Secondary structure assigments	III
	Ligand docking	I
X-ray Crystallography	Crystallisation	II
	Molecular replacement	II
	Model building and refinement	II
	Structural alignment	II, III

4.2 Crystallisation of recombinant human GAA

The recombinant human GAA (rhGAA, Myozyme) is a 110 kDa precursor form of the enzyme, which is used in enzyme replacement therapy of Pompe patients. It was a gift from Genzyme corporation and was shipped to us in a lyophilised form. Upon resolubilisation with water, rhGAA had a concentration of 5 mg/ml and was in a protein buffer containing 25 mM sodium phosphate pH 6.2, 2 % mannitol, 0.005 % polysorbate 80.

The crystallisation conditions of rhGAA were screened using the automated protein crystallisation facility at the Institute of Biotechnology (http:// www.biocenter.helsinki.fi/bi/xray/automation/). In some crystallisation trials, the resolubilised rhGAA was concentrated from the original 5 mg/ml to 15 mg/ml and was in others exchanged into 0.02 M sodium acetate pH 4.3. In all crystallisation trials, the protein was filtered through a 0.22 µm filter (Millipore) prior to crystallisation. The best diffracting crystals grew from a reservoir solution containing 0.1 M Tris pH 7, 11 % glycerol, 28% PEG 600 and 0.5 M ammonium sulphate with the protein solution containing 7 mg/ml rhGAA in 0.02 M sodium acetate pH 4.3. Crystals were flash cryocooled in cryoloops (Molecular dimensions) for data collection.

5 Results and discussion

5.1 bPLBD1 is a lysosomal Ntn-hydrolase (Study II)

Previous proteomic studies suggested that phospholipase B-like protein 1 is a putative lysosomal hydrolase (Sleat *et al*., 2005; Sleat *et al*., 2008; Della Valle *et al*., 2011; Chen *et al*., 2009). We investigated the subcellular localisation of PLBD1 by confocal microscopy. bPLBD1 overexpressed in CHO and HeLa cells extensively co-localised with the lysosomal marker LAMP1 in vacuolar structures. This further suggests that PLBD1 is a lysosomal protein.

5.1.1 Processing and glycosylation of bPLBD1

I purified PLBD1 from bovine kidneys. The advantage of isolating protein from an endogenous source was that I could characterise the post-translational modifications of the mature bPLBD1 reflecting the situation *in vivo.* The characterisation showed that the 545 residue long bPLBD1 precursor is proteolytically processed and N-glycosylated for achieving the mature protein. On a reducing SDS-PAGE experiment, bPLBD1 formed three major bands with molecular weights of approximately 45 kDa, 22 kDa and 11 kDa (Figure 6A). This together with N-terminal sequencing suggested that, in addition to the removal of the signal sequence, bPLBD1 is proteolytically cleaved in two places, resulting in the removal of about 19 residues between the N- and C-terminal segments (Figure 6C). These results are similar to those obtained with the human neutrophile derived PLBD1, which also forms two bands (42 kDa and 22 kDa) on SDS-PAGE (Xu *et al*., 2009), but disagree in the placement of the proteolytic cuts. In bPLBD1 the cuts are only between the two segments, not in both to the N-and C-terminus of the 42 kDa band, as proposed for the human PLBD1.

I did time resolved deglycosylation, where the N-glycans are removed stepwise with PNGase F to study if all N-glycosylation sites in mature bPLBD1 are utilised. PNGase F is capable of removing all types of N-glycans and the experiment showed that four N-glycans from the C-terminal segment and one from the N-terminal segment were removed (Figure 6B). Thus all five glycosylation sites present in the mature bPLBD1 sequence are in use.

I studied the nature of the glycans by deglycosylating bPLBD1 with Endoglucosidases F1, F2 and F3 and Endo H_f and PNGase F. Endoglucosidases F2

and F3 cleave only complex glycans and failed to release any glycans from bPLBD1. On the contrary, the high mannose or hybrid type glycan hydrolysing Endoglucosidase F1 (native samples) and Endo H_f (denatured samples) cleaved all glycans from bPLBD1 and resulted in a similar molecular weight bPLBD1 peptide as with PNGase F treatment (Figure 6A). This suggests that all the N-glycans in PLBD1 are of the non-complex type.

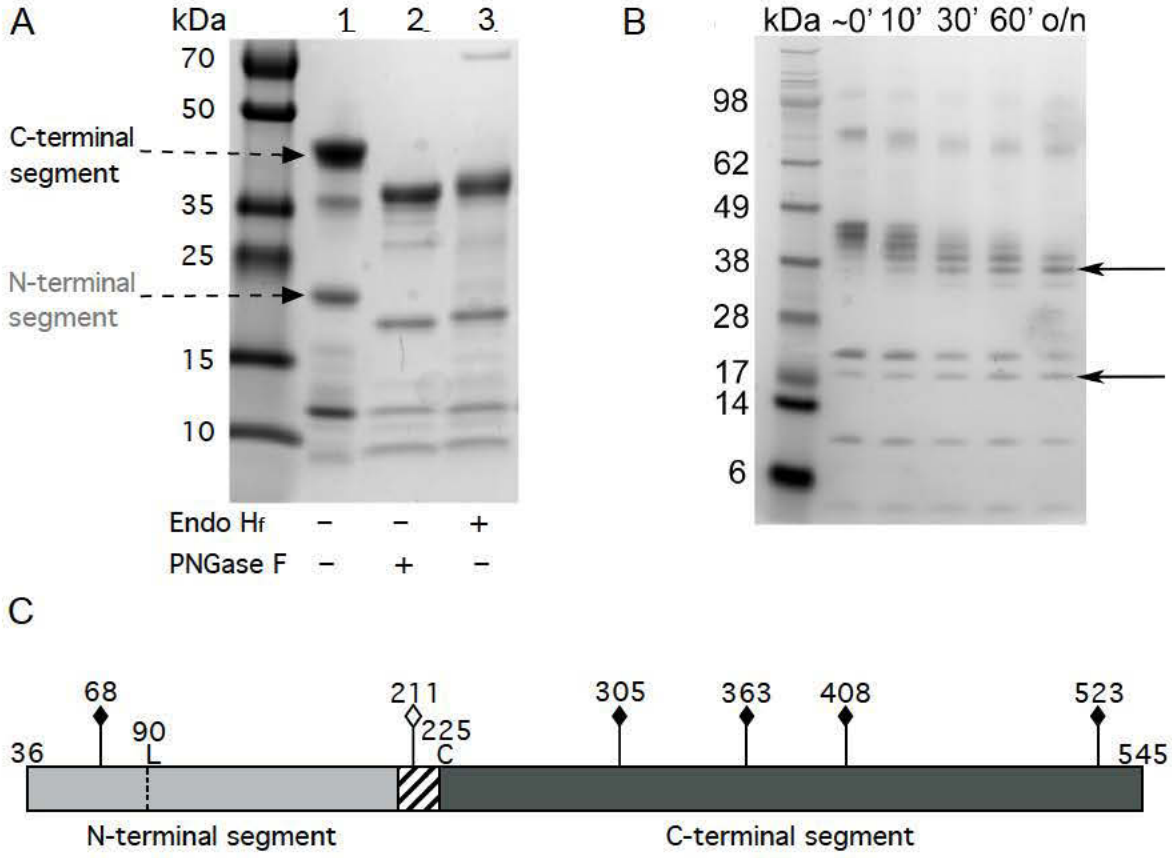


Figure 6. Primary structure and glycosylation of bPLBD1. *A) The molecular weight standard Page Ruler Prestained Protein Ladder (Thermo Scientific) is on the left; lane one has untreated bPLBD1; lanes two and three completely deglycosylated bPLBD1. The N- and C-terminal segments are indicated with an arrow B) Time resolved deglycosylation. Molecular weight standard Blue Plus2 Pre-Stained Standard (Invitrogen) is on the left. Deglycosylation samples are immediately after mixing (~0'), 10 minutes (10'), 30 minutes (30'), 60 minutes (60') and overnight (o/n) incubation. The arrows point to the fully deglycosylated peptides. C) Experimentally verified N-glycosylation sites are marked with solid diamonds on the primary structure. The ~19 residue long dashed peptide is absent from the mature protein and thus the use of the N211 site is unverified and marked with a hollow diamond. The SDS-PAGE was stained with Bio-Safe Coomassie stain (Biorad).*

5.1.2 bPLBD1 structure

I crystallised the kidney purified bPLBD1 and solved its X-ray structure at 1.9 Å resolution to an R_{work} of 19.0 % and an R_{free} of 23.1 % by using molecular replacement and the mPLBD2 structure (pdb 3fbx; Lakomek *et al*., 2009) as a template. The structure containing 491 residues verified the predicted proteolytic processing, as there is no electron density for the residues which are proteolytically removed between the N- and C-terminal segments. There was one complete bPLBD1 molecule per asymmetric unit. Instead of forming separate domains, the segments together create the αββα-fold of the molecule (Figure 7A). The electron density was visible at least for the first GlcNAc residue of all five N-glycans (Figure 7A) confirming the N-glycosylation pattern from the deglycosylation studies. The mature bPLBD1 sequence has 5 cysteine residues and four of them form two disulphide bridges, C467-C472 and C471-C468, in the C-terminal segment (Figures 7A and 7C).

The molecular mass of bPLBD1 estimated based on the gel filtration elution volume was 130 kDa, which is significantly larger than the calculated molecular mass of the mature protein, 58.8 kDa. If the protein were a dimer with five average size N-glycans in both monomers, the calculated molecular mass would correspond to that estimated from gel filtration. By using the crystal symmetry matrices to generate symmetry-related copies of the bPLBD1 molecule, I was able to identify a monomer-monomer assembly that has a large enough interface (1133 $Å^2$), to be considered a true dimer-interface (Figure 7B). For an area, on the surface of a protein, the interface is notably conserved in the PLBD1 family, whereas the corresponding region in the monomeric PLBD2 family has very little conservation. Thus, the biological unit of bPLBD1 is a dimer.

5.1.3 The Ntn-fold and substrate binding Site

The monomer of bPLBD1 and mPLBD2 have the N-terminal nucleophile aminohydrolases (Ntn-hydrolase) superfamily fold (Figure 7A). In Ntn-hydrolases, proteolytic cleavage is needed to release the N-terminal amino acid, which in the case of PLBD1 is Cys225 (Figure 7A). The N-terminal nucleophile has two catalytic roles: as a nucleophile and as a general base (Oinonen and Rouvinen, 2000). The surroundings of Cys225 in bPLBD1 are very conserved, which is consistent with the fact that the active site of Ntn-hydrolases is located around the N-terminal nucleophile (Figures 7A and C). From the bPLBD1 structure, it is evident that the removal of a peptide approximately 19 residues in length between

the N- and C-terminal segments is necessary for the opening of the active site for substrate binding (Figure 7A).

Among the Ntn-hydrolases with known function, the most similar structure to bPLBD1 was that of acyl coenzyme A isopenicillin N acyltransferase from

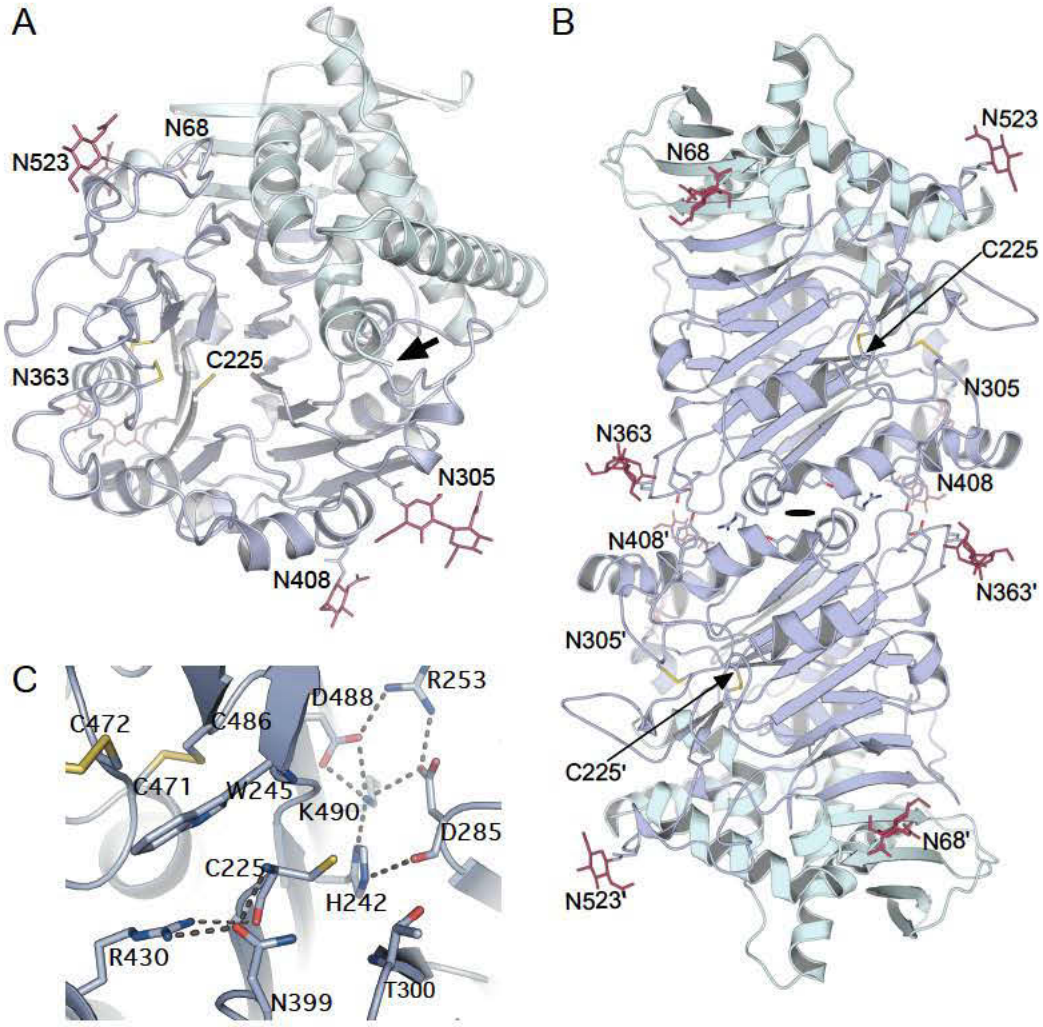


Figure 7. Cartoon representation of the bPLBD1 structure. *The N-terminal segment is shown in turquoise and the C-terminal segment in light purple. N-linked GlcNAc molecules that were visible in the density are shown as red sticks as are the asparagine residues they are linked to. Disulphide bridges and the C225 is shown in yellow. A) The monomer consists of a four layer αββα sandwich core, which is a typical Ntn-hydrolase fold. The active site is opened by a autocatalytic cleavage at C225 and by another cleavage indicated by an arrow. B) The dimer rotated by about 90° from the view in A. C225 is indicated with an arrow. The two-fold axis is perpendicular to the plane of the paper. C) The residues surrounding the N-terminal nucleophile C225. The catalytic machinery is supported by H242 and the salt bridge network that extends towards R253.*

Penicillium chrysogeum (acyltransferase, pdb 2X1E; Bokhove *et al*., 2010). Ntn-hydrolases have great similarities in their active sites. This allowed identification of active site amino acid residues in addition to the C225 that are important for catalysis. Typically the oxyanion hole of the Ntn-hydrolase is formed by a backbone amide and an asparagine Nδ atom (Bokhove *et al*., 2010). The oxyanion hole residues in acyltransferase superimposed on Thr300 and Asn399 in bPLBD1. In the acyltransferase, the Asp121 backbone amide stabilises the thiolate state and the side chain oxygen the positively charged α-amino group of the N-terminal cysteine (Bokhove *et al*., 2010). In bPLBD1, Trp245 is located in the same position (Figure 7C).

5.1.4 Is the PLBD1 an amidase?

Despite the structural similarity to other Ntn-hydrolases, I could not deduce the bPLBD1 substrate by active site comparison. The reason for this is that even though key residues are well conserved, the active site machinery of Ntn-hydrolases is so versatile that it can accomodate a variety of different substrates and even the acyltransferase is evolutionarily too distant to bPLBD1. Nevertheless, there is a homologous protein to PLBD1 for which the function had been previously assigned: the *Dictyostelium dictyostelium* homolog had been shown to have Phospholipase B (PLB) activity (Morgan *et al*., 2004). The other PLBD1 and PLBD2 members have also been proposed to be phospholipases based on the ~30% sequence identity to the *Dictyostelium* protein. After prolonged storage, very low PLB activity has also been detected for PLBD1 purified from human neutrophils (Xu *et al*., 2009), but not for any other family member (Deuschl *et al*., 2006; Jensen *et al*., 2007).

Based on the structural analysis, I believe that despite the detected PLB activities, these proteins are not phospholipases, because PLB activity is inconsistent with the protein structures. First of all, the bonds that are hydrolysed in all of the known Ntn-hydrolase substrates are amide, not ester bonds. The esterase function can not be excluded, because ester bonds are easier to break than amide bonds. Second, in terms of a "perfect catalyst"(Albery and Knowles, 1977), the mechanism of Ntn-hydrolases, which is not present in esterases or other amidases, achieves maximum lowering of the free energy of the difficult step in amide hydrolysis, the activation of the leaving group. This would be wasted on ester bonds. In Ntn-hydrolases, the incipient primary RNH2 leaving group is stabilised by the equivalent primary amine group of the N-terminal nucleophile and such stabilisation is not necessary for esterases. Third, even if the PLBD1 substrate

would contain an ester bond, phospholipids are not suitable putative substrates for PLBD1 because the binding site is incorrect. The orientation of the ester bond, if it were the same as the amide, would place the long hydrophobic tail of the phospholipid in what in bPLBD1 is a small cavity, about the size of a arginine (Study II Figure 7). In mPLBD2, this acyl group binding cavity is slightly larger, but the structural analysis shows that in neither of the proteins is it sterically possible to place the phospholipid tail in this cavity. As the PLB activities shown by Morgan *et al* (2004) is for partially purified protein and by Xu *et al* (2009) only after prolonged storage, it is possible that this is merely a side reaction for these enzymes or due to some impurity.

The substrates for PLBD1 and PLBD2 proteins are probably somewhat similar, but not the same, because they have such large differences in their substrate binding sites. As mentioned above, the acyl group binding cavity in mPLBD2 is larger than in bPLBD1. Secondly, the other part of the site is in both proteins a fairly large cleft, but in mPLBD2, it is partly covered by two loops. The shape and the size of the open cleft of the substrate binding site in bPLBD1 suggest that the putative substrate could be a sugar based compound, such as sialic acid, since it has a small enough acyl group to fit into the small cavity.

5.2 Lysosomal proteins have several common features (Studies I, II, and III)

The lysosomal lumen creates both advantages and great challenges for soluble lysosomal hydrolases. The acidic environment in itself facilitates hydrolysis of macromolecules and in addition lysosomal hydrolases have optimal enzymatic activity at low pH. On the other hand, the acidic environment and the presence of hydrolytic enzymes is challenging for all proteins, and the lysosomal enzymes are at constant risk to degrade themselves. The lysosomal lumen thus presents a unique environment inside the eukaryotic cell and lysosomal hydrolases have unique features to cope with this.

Soluble lysosomal hydrolases are a group of mostly unrelated, structurally and functionally diverse proteins. We chose 39 lysosomal protein structures (Table 4) for our structural analysis so that they would form a representative set of soluble lysosomal proteins (Study III). Selection was done by choosing all mammalian sequences with structural information marked as lysosomal in UniProt (Consortium, 2012), limiting to those that also including signal sequence and

Table 4. Fold architecture on soluble lysosomal proteins. Table omitted from Study III.

CATH Architecture (number)	*Protein*	*Salt Bridges*	*Disul-phides*	*Proline Clusters* PP	PXP
Orthogonal bundle (1.10)	Proactivator polypeptide	5	4	-	2
	Myeloperoxidase	35	6	2	3
β-*barrel (2.40)*	Cathepsin D	11	4	1 (PPP)	-
	Cathepsin E*	6	3	2	-
Sandwich (2.60)	Epididymal secretory protein E1	1	3	-	3
Distorted sandwich (2.70)	Ganglioside GM2 activator	4	4	-	2
α–β *barrel (3.20)*	Hyaluronidase-1*	16	5	-	2
3-layer(αβα) sandwich (3.40)	Lysosomal thioesterase PPT2	10	3	-	-
	Palmitoyl-protein thioesterase 1	16	3	2	3
	Gamma glutamyl hydrolase	8	-	-	-
	Lysosomal protective protein	17	4	1	-
	Dipeptidyl peptidase 2 *	18	4	-	-
	Lysosomal pro-x carboxypeptidase*	21	4	-	2
	Tripeptidyl-peptidase 1*	29	3	4	5
	Proprotein convertase subtilisin *	17	12	3	1
4-layer sandwich(3.60)	N(4)-(β-*N*-acetylglucosaminyl)-L-asparaginase*	15	4	1	3
	Putative phospholipase B-like 1				
	Putative phospholipase B-like 2*	31	2	2	2
	Tartarate resistant acid phosphatase type 5	12	1	-	1
α–β *complex (3.90)*	Cathepsin Z	18	6	-	-
	Cathepsin K	13	3	-	-
	Cathepsin S	23	3	-	-
	Cathepsin L1	16	3	-	-
	Cathepsin H	9	4	1	1
	Cathepsin B	21	6	3	1
	Cathepsin F	6	3	-	-
2-Domains					
α–β *barrel & sandwich (or 2-layer sandwich)*	α-*N*-acetyl galactosaminidase	25	4	2	1
	α -galactosidase A	30	5	-	-
	β -hexosaminidase subunit α	32	3	-	1
	β -hexosaminidase subunit β	35	3	-	1
	Glucosylceramidase	23	2	1	-
	Galactocerebrosidase*	16	1	PPPP	2
	β galactosidase*	32	2	1	4
3-layer(αβα) sandwich & 2-layer sandwich	Arylsulphatase A	32	6	5	2
	Arylsulphatase B	36	4	3+PPP	1
	N-acetylgalactosamine-6-sulphatase	41	3	4	1
α–β *complex &* β-*barrel*	Dipeptidyl peptidase 1 (Cathepsin C)	16	5	-	1
3-Domains					
α–β *barrel & 2x sandwich*	β glucuronidase	44	-	2	2
5-Domains					
α-β *barrel, up-down bundle, sandwich, distorted sandwich & sandwich*	Lysosomal α –mannosidase*	48	4	3	1

verifying with the help of literature (Sleat *et al*., 2008; Schröder *et al*., 2010) that they were part of lysosomal proteome. In addition, one mammalian structure was selected within structures of 70% sequence identity and if a single protein had several structures the full length mature human protein was preferred.

The protein domains of the representative set of soluble lysosomal proteins belong to three most common fold classes: mainly α, mainly β, or α-β. The most common domain architecture (8 out of 39 proteins) is the 3-layer αβα-sandwich (Table 4).

5.2.1 Thermal stability of lysosomal proteins reflects stability against lysosomal conditions (Studies I, II and III)

Lysosomal proteins show often good thermostability. I measured the pH and thermal stability of recombinant human acid α-glucosidase (rhGAA) and bPLBD1 by differential scanning fluorimetry, which determines the midpoint of thermal unfolding (T_m). Both proteins have relatively high T_ms. The highest T_m for rhGAA was 70.2 °C at pH 7, and for bPLBD1 74.5 °C at pH 5.0 (Figure 8). Similarly high T_m values have been measured previously for other lysosomal proteins. For example the T_m of human acid α-galactosidase (GLA) is 60.2 °C at pH 5.2 (Lieberman *et al*., 2009) and lysosomal α-mannosidase (MAN2B1), which purifies together with PLBD1 tolerates temperatures of at least 60°C (Tollersrud *et al*., 1997). However, unlike GLA and MAN2B1, which have narrow pH-optimum T_m, rhGAA and bPLBD1 are stable over a wide pH range; from pH 4 to over 7 (Figure 8).

Proteins are usually only marginally stable and tend to denature close to the growth temperatures of the organism (Richter *et al.*, 2010). Therefore, the T_ms of these lysosomal proteins are significantly higher than what would be expected for mammalian proteins.

The basis for thermostability of proteins from thermophiles, micro-organisms or viruses that thrive in high temperatures, has been extensively studied (Szilagyi and Zavodszky, 2000; Gromiha *et al*., 2013). Amino acid distribution, higher number of salt bridges and hydrogen bonds (Kumar *et al*., 2000) and oligomerisation (Taka *et al*., 2005) contribute to the stability of these enzymes.

It is not clear to what extent the same factors contribute to lysosomal protein stability. Thermostable species are evolved to function in extreme temperatures, such as the optimum temperature of *Thermotoga maritima*, 80°C (Huber *et al*., 1986). Thermostable proteins are in general more rigid compared to the lysosomal

hydrolases, which require conformational flexibility due to their 37 °C working environment. What is the structural basis of the stability of proteins in lysosomal lumen?

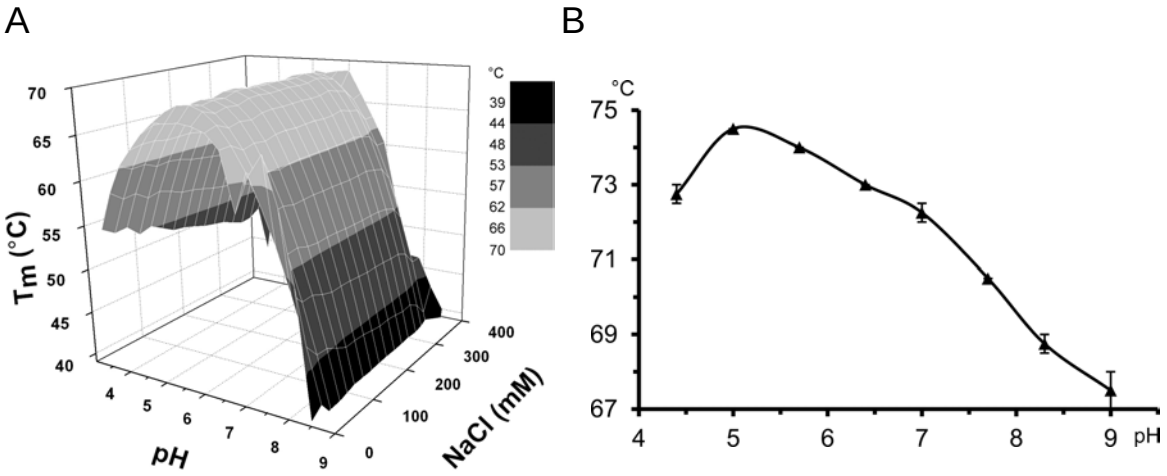


Figure 8. The pH dependence of the thermal denaturation of rhGAA and bPLBD1. *A) Stability surface for rhGAA shows T_m plotted as a function of NaCl concentration and pH. The following buffers, which were selected to get a representative set of the results, are presented on the surface: acetate (pH 3.5, 4, 4.5, 5 and 5.5), MES (pH 6 and 6.5), Hepes (pH 7, 7.5 and 8) and borate (pH 8.5) with 0, 40, 80, 160, 240, 320 and 400 mM NaCl. B) The pH dependence of bPLBD1 thermal denaturation was measured in a mixture of Succinic acid, Sodium Phosphate and Glycine buffers to avoid chemical effect due to changing the buffer during titration. Standard error bars are displayed for each data point.*

In the set of 39 representative lysosomal protein structures analysed in this study (III), only 12 proteins have higher than expected proline content. Thus, unlike in many thermostable proteins, high overall proline content does not appear to be a common feature of lysosomal proteins. However, multiple PP or PXP patterns, which are located in adjacent regions of the protein sequence and often on surface exposed loops, are common in lysosomal proteins (Table 4). Prolines stabilise loop regions by restricting the number of possible conformations and protect the loop regions from proteolysis.

Some lysosomal hydrolases such as MAN2B1 have extensive salt bridge networks (Heikinheimo *et al*., 2003). However, a large number of salt bridges is not characteristic for lysosomal enzymes. There was no significant difference in the number of salt bridges or charged residues in the representative set of protein structures compared to human proteins in general.

Lysosomal proteins have features that are beneficial for protein stability, but which do not occur in thermophilic bacterial and archeal proteins. The two most prominent features are N-glycosylation and disulphide bridges. The N-glycans have an important role for lysosomal hydrolases in lysosomal targeting *via* the M6P-dependent route. In addition, N-glycans stabilise protein structure (Wormald and Dwek, 1999; Hanson *et al*., 2009) and protect them from proteases. All the analysed 39 lysosomal hydrolases have at least one N-glycan. The number of N-glycans vary within the set, and the number of glycosylation sites does not correlate to the length of the polypeptide (Study III, Table I).

Disulphide bridges stabilise the protein fold by adding a covalent link between different secondary structures (Matsumura *et al*., 1989). All proteins in the selected set, except γ-glutamyl hydrolase, contain disulphide bridges and most proteins contain more than one (Table 3). Like the multiple proline patterns, disulphide bridges are typically on surface areas containing long loop structures. In PLBD1 and GLA, disulphide bridges support the wall of the substrate binding cleft.

Somewhat surprisingly, although disulphide bridges are common in lysosomal proteins, their number and position are not well conserved in similar folds, not even between homologous proteins. For example, the disulphide pattern of PLBD1 and PLBD2 are not conserved, despite the highly similar monomer structure. Similarly, the disulphide bridges in sulphatases are not conserved in all three lysosomal sulphatases. Cysteine cathepsins with a similar fold have three to six disulphide bridges, of which two are completely conserved. The non-conserved disulphide bridges may provide additional protection against proteolytic turnover. For example, in the region where cathepsin C lacks one disulphide bridge that cathepsins K, S, L1 and F have, cathepsin C has an additional proteolytic cleavage site.

Thermophilic proteins use combinations of several different mechanisms to achieve the same stability (Szilagyi and Zavodszky, 2000). Even related proteins can use different mechanisms (Repo *et al*., 2012). Our work indicates that this is

true also for lysosomal proteins. Only a few structural features providing stability are shared by all or even related lysosomal proteins.

5.2.2 M6P-dependent transport signal (Studies II and III)

The intriguing enigma of the M6P-dependent transport signal arises from the following. The GlcNAc phosphotransferase recognises large number (over 50) of the newly synthesised soluble lysosomal proteins, which however lack a universal and evolutionary conserved structural recognition motif. Nonetheless, recognition by the GlcNAc phosphotransferase is specific only to lysosomal proteins and to a few secreted proteins (Reitman and Kornfeld, 1981; Dahms *et al*., 2008).

5.2.2.1 Conservation of the lysosomal transport signals (Study III)

I first analysed the conservation of the N-glycosylation sites in the four lysosomal proteins, which have been experimentally verified to contain M6P and where the recognition signal for GlcNAc phosphotransferase is known. Phosphorylated N-glycosylation sites in cathepsin D (Dittmer *et al*., 1997) are conserved in 22 (N134) and in 23 (N263) of the 23 known mammalian sequences; in cathepsin L N221 (Cuozzo *et al*., 1998) is found in all 12 known sequences; in AGA both sites (N38 and N308) (Tikkanen *et al*., 1995) are conserved in 20/22 mammalian sequences; and in arylsulfatase A both sites (N158 and N350) (Yaghootfam *et al*., 2003) are conserved in all 20 known mammalian sequences. Thus, the site of the phosphorylated glycan seems to be well preserved.

I then similarly studied the conservation of the lysine residues known to affect the N-glycan phosphorylation levels in these proteins. In arylsulfatase A, the only lysine residue known to affect phosphorylation, K457 (Yaghootfam *et al*., 2003), is conserved in all 20 known mammalian orthologous sequences. However, six other lysine residues in arylsulfatase A sequence are equally well conserved. In contrast, in cathepsin D two out of the three most important lysine residues are well conserved (K267 in 20/21 and K331 in 21/21) but the third (K357) only in 12 sequences. K267 is the most important lysine to modulate phosphorylation of cathepsin D conserved glycan at N263 (Steet *et al*., 2005). In cathepsin L, which is not related to cathepsin D, all five lysine residues that affect phosphorylation are conserved in 11 or 12 of the 12 known sequences. The AGA lysine residues (K177, K183, K214) are conserved in all or almost all sequences.

Are the well conserved phosphorylated glycosylation sites or lysines also conserved in other proteins with similar structure? K457 in arylsulfatase A is also conserved in the paralogous arylsulfatase B, but has no significance on

arylsulfatase B phosphorylation (Yaghootfam *et al*., 2003). The location of the N-glycans in the overall fold is similar in these two sulphatases, but the glycosylation sites are not conserved in sequence. This lack of N-glycosylation conservation might explain why K457 is not important for arylsulfatase B phosphorylation. Comparison of cathepsin L structure to the other cysteine cathepsin structures reveals that the N-glycosylation and surface lysine locations are conserved only in the closest relative in the protein family, cathepsin V. Cathepsin L and V have ~70 % sequence identity and they share N-glycosylation at N221. In addition, cathepsin V has another glycosylation site. K54 and K99, which are part of the cathepsin L GlcNAc phosphotransferase signal, are present in both proteins. Unfortunately, there is not yet evidence of the significance of these lysines for cathepsin V lysosomal localisation.

Some elements of the lysosomal transport signal, such as the phosphorylated N-glycans, are well conserved within mammalian sequences of an individual protein. Lysine residues, although on the generally less conserved protein surface, are also surprisingly well conserved. Most lysines are conserved in almost all mammalian sequences and thus the lysine residues that are part of the recognition signal, do not substantially stand out. Less-conserved lysines, which are part of the recognition signal, usually contribute only to the phosphorylation level, but are not part of the minimally required recognition pattern. The recognition signal is conserved only in very closely related proteins, and not necessarily in other proteins with a similar fold, as shown by the cysteine cathepsin family, where only the very similar cathepsins L and V share all the components of the recognition signal.

The lysine residues included in the recognition pattern are located in varying secondary structure elements, such as in a β-hairpin, an α-helix, a β-strand, in a disulphide supported loop and even in the propeptide region. Apart from the propeptide lysine in cathepsin L, all important lysines are located on non-flexible, well supported areas. One of the lysine residues is typically in close proximity to the phosphorylated glycan, but I found no universal distances between the lysine residues and the (phosphorylated) N-glycan, as previously suggested (Cuozzo *et al*., 1998; Warner *et al*., 2002). Arylsulphatase A is an exception because only one lysine residue is known to affect binding of the GlcNAc phosphotransferase (Yaghootfam *et al*., 2003).

In cathepsin L, it was shown that several glycans can be phosphorylated with GlcNAc phosphotransferase binding at only one site because of the N-glycan

flexibility (Warner *et al*., 2002). This would explain why precise lysine to N-glycan distances are not necessary on the protein surface. It also seems evident that proteins with several phosphorylated N-glycans that are located on distant parts of the protein structure must have several GlcNAc phosphotransferase recognition motifs. For example, in cathepsin D, the known recognition signal affects phosphorylation of only one of the N-glycans (Steet *et al*., 2005). In this respect it is understandable that proteins with similar structure, but no conserved glycosylation, do not share conserved lysines for GlcNAc phosphotransferase recognition.

5.2.2.2 Putative transport signal of the phospholipase B-like proteins 1 and 2 (Study II)

One of my main interests in the the bPLBD1 structure was to identify surface features on bPLBD1 that might be part of the GlcNAc phosphotransferase recognition site. bPLBD1 most likely follows the M6P- dependent pathway as it was shown in this study to carry only non-complex type glycans and was previously shown to contain M6P-glycan (Sleat *et al*., 2008). Based on the protein structure bPLBD1 has such a large number (32) of surface exposed lysine residues, that it prevents identification the important lysine and N-glycan residues by analysing the bPLBD1 structure alone. However, the existence of the paralogous protein structure, mPLBD2 (PDB 3FBX; Lakomek *et al*., 2009), offered a chance to identify the most conserved residues.

bPLBD1 and mPLBD2 have one N-glycan, N408 (N441 in mPLBD2), which is conserved in both families in sequence and structure (Figure 9A). Thus they could share lysine residues important for recognition by the GlcNAc phosphotransferase. In addition, the N-glycans that are not conserved between the two PLBDs have a similar distribution on the protein surface (Figure 9). Some glycosylation sites (N305 in bPLBD1 and N236 in mPLBD2) cluster close to the conserved N408 and could be phosphorylated with the same GlcNAc phosphotransferase binding area.

bPLBD1 and mPLBD2 have only one surface lysine in common (K358 in bPLBD1). This is located 33-37 Å away from the conserved N408(N441) glycosylation site and is a good candidate for the distal lysine of the recognition signal. As there are no lysine residues conserved between the families, I then searched for lysine residues that are conserved within the families and located in the structural region between K358 and N408. There were two such lysine residues in each protein, K342 and K334 in bPLBD1, and K372 and K418 in mPLBD2. All four lysine sites

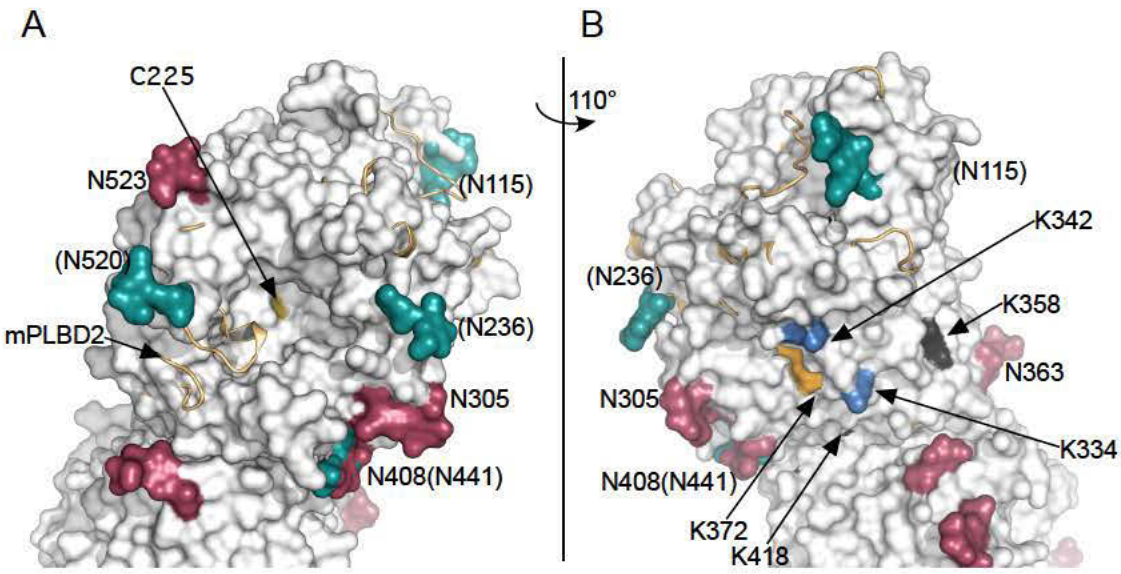


Figure 9. N-glycosylation and GlcNAc phosphotransferase recognition signal of bPLBD1 and mPLBD2. *bPLBD1 is shown as white surface representation with N-glycans and the asparagines they are attached to as red and mPLBD1 as yellow cartoon with the N-glycans and the asparagines they are attached to as turquoise.Only one monomer of bPLBD1 is shown in its entirety. The mPLBD1 cartoon is only partly visible because it is embedded in the bPLBD1 surface. A) The orientation of PLBD1 is as in figure 6. The key conserved N408 (N441 in mPLBD2) is to the bottom left. The arrow shows the location of the C225 nucleophile in the active site. B) Rotated by 110° from A). The arrows indicate the lysine residues that are probably an important part of the GlcNAc phosphotransferase recognition site when it phosphorylates N408 (N441).*

(K334, K342 in bPLB1 and K372, K418 in mPLBD1) are well conserved in their protein families (Table 5). K342 (bPLBD1) and K372 (mPLBD2) are not analogous but are located in the same structural element and so could be the proximal lysine residues of the recognition motif. However, these findings need to be confirmed by experimental studies.

5.2.3 Spectrum of LSD associated missense mutations (Study III)

In order to understand the molecular mechanism on how the disease associated mutations lead to an outbreak of a lysosomal storage disorder, I chose four lysosomal proteins, which have a known or reliably modelled three dimensional structure. These proteins were *N*-acetylgalactosamine-6-sulfatase (GALNS) (pdb 4FDI, Rivera-Colón *et al.*, 2012), human α-galactosidase (GLA) (pdb 3HG3, Guce

Table 5. Conservation of the N-glycosylation and surface lysine residues in the mammalian PLBD1 and 2 families.

bPLBD1	Conservation in all known[a] mammalian sequences	mPLBD2	Conservation in all known[a] mammalian sequences
N68	22/22	N93[b]	17/18[c]
		N115	19/22
N211[d]	13/22	N236	22/22
N305	22/22		
N363	15/22		
N408	20/22	N441	22/22
N523	22/22	N520	21/21[e]
K334	16/22	K418	20/22
K342	16/22	K372	21/22
K358	22/22	K389	20/22

[a]Jan 29th, 2013
[b]glycan not built in the structure (PDB ID: 3FBX)
[c]five of the sequences start after this residue
[d]glycan not visible in the structure due to proteolytic removal, so it is unclear if this site is used or not.
[e]one sequence ends before this residue

et al., 2010), human lysosomal α-mannosidase (MAN2B1) (Kuokkanen *et al.*, 2011) and lysosomal α-glucosidase (GAA), (modelled based on intestinal maltase-glucoamylase in study I). Each protein has a different amount of known disease-associated missense mutations; 386 Fabry disease mutations in 209 sites (54.1 % of sites in sequence) are known for GLA (Saito *et al.*, 2011), 188 Pompe disease mutations in 144 sites (15.6 % of sites) are known for GAA (www.pompecenter.nl) and 139 Morquio A mutations in 118 sites (24 % of sites) in GALNS are disease associated (Rivera-Colón *et al.*, 2012; Morrone *et al.*, 2014), but for MAN2B1 only 41 α-mannosidosis disease mutations in 36 sites (4% of sites) are known (www.amamutdb.no). The tolerated changes; 21 sites in GAA, 9 sites in α-mannosidosis, three sites in Fabry and two in GALNS, form an important control for the analysis.

5.2.3.1 Distribution of the disease-associated mutations among the 20 amino acids

The disease associated missense mutations were unevenly distributed among the 20 amino acids. The relative frequency of the mutations did not reflect the abundance of the individual amino acid in the protein sequence (Figure 10A).

Most frequently the disease associated mutations affected glycine, arginine, leucine, proline and aspartate residues (Figure 10A). Leucine occurs among mutations less frequently than expected based on the sequence abundance in all but the MAN2B1 enzyme (Study III, Supplementary Figure I). Leucine is most often (in 35 out of 69 mutations) replaced by proline. In addition to leucine, disease-associated mutations occur less frequently than expected at other small and core forming hydrophobic residues, such as alanine, valine, isoleucine and phenylalanine (Figure 10A).

Contrary to leucine, glycine and arginine associate with a disease much more often than expected based on their abundance in sequences (Figure 10A, Study III, Supplementary Figure I). In three of the four cases, glycine is the most frequently mutated site (Supplementary Figure I). In addition, glycine has several sites that contain more than one type of mutation (Figure 10B). In about one third of the mutations glycine residue is replaced by arginine.

Although arginine is not a common residue in protein sequences, arginine disease-associated mutations are common and clearly more frequent than would be expected in all four proteins (Figure 10A, Study III Supplementary Figure I). Typically, arginine sites also contain more than one type of mutation (Figure 10B). Arginine residues are replaced by cysteine, histidine, glutamine, proline and tryptophan with almost equal frequency in about ten of the 69 arginine mutations.

In addition to glycine and arginine, some low abundance amino acids, such as tryptophan, cysteine and methionine, occur more frequently than expected among disease associated mutations.

Proline and cysteine disease-associated mutations are more common than expected in all but GALNS (Study III Supplementary Figure 3) and the proline mutations are among the three most common mutations in two of the four proteins (Study III Supplementary Figure 3). Proline is most often replaced by leucine. Cysteine mutations are especially common in GLA (Study III Supplementary Figure 3).

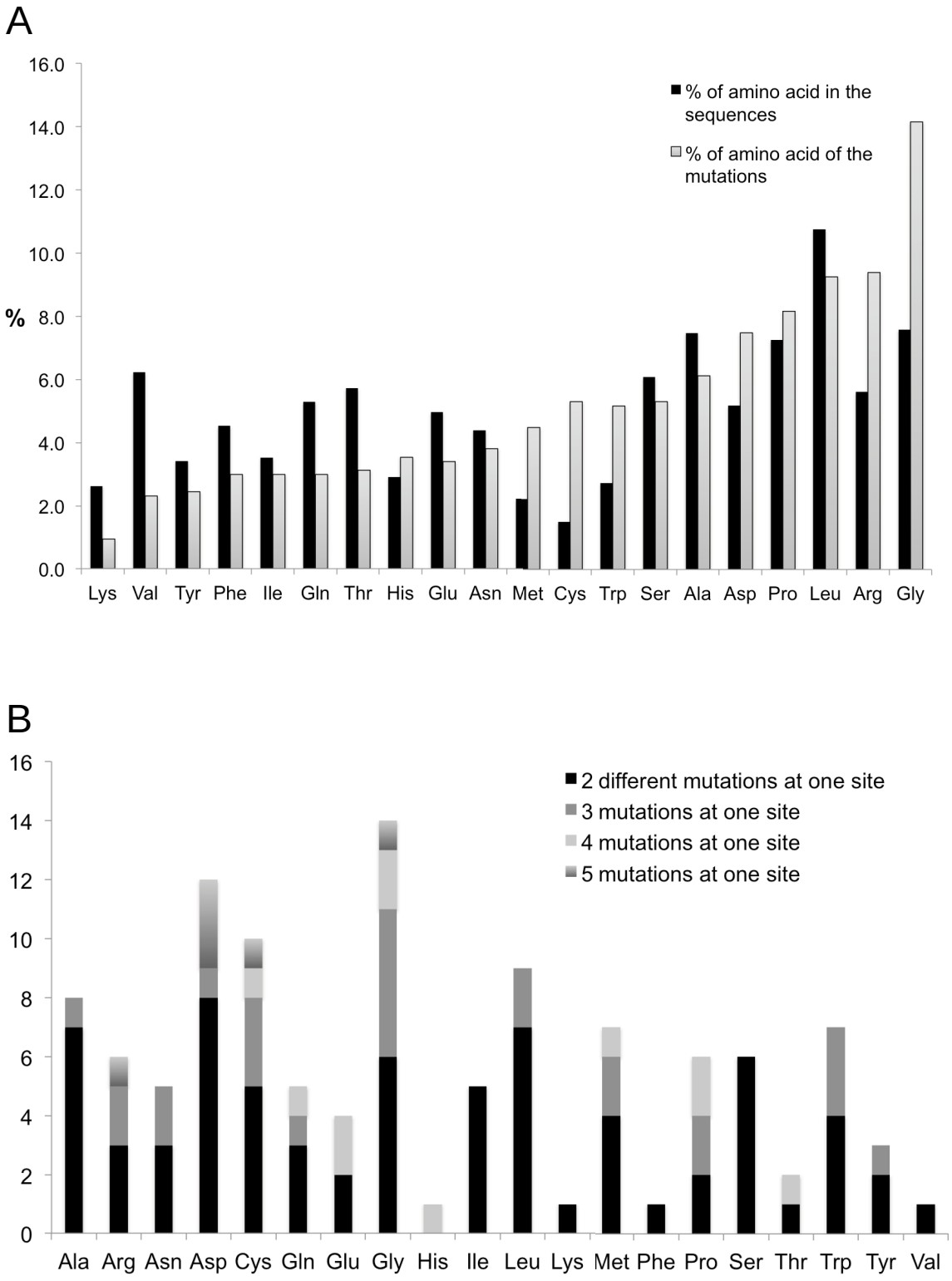


Figure 10. The distribution of LSD-associated mutations. *A)The overall distribution of amino acid types in the four lysosomal protein sequences (black bars) and in disease-associated mutations (grey bars). B) The number of residues that contain more than one type of different disease associated mutations. The analysed proteins were α-galactosidase A (GLA), lysosomal α-glucosidase (GAA), N-acetylgalactosamine-6-sulphatase (GALNS) and lysosomal α-mannosidase (MAN2B1).*

Aspartic acid and histidine disease-associated mutations show great variation in how they appear in the list of the disease-associated mutations in the individual proteins. Aspartic acid mutations occur frequently only in GLA, where they are not only over-represented but also the second most common type of mutation (Study III Supplementary Figure 3). In addition, aspartic acid sites in GLA are mutated to many different types of residues (Study III Supplementary Figure 2). Similarly, histidine mutations are clearly over-represented in MAN2B1, but not in other proteins (Study III Supplementary Figure 3).

5.2.3.2 Structural analysis of the most common mutation types

I analysed the five most common types of disease-associated missense mutations in protein structures, in order to understand how they interfere with the protein function.

The ratio between surface exposed and buried residues varies from one residue to the next (Table 6). Residues containing disease associated mutations are more often among buried residues than the residues of that amino acid in general (Table 6). This is especially evident in arginine and glycine residues, which in general are found more often on the protein surface, but the residues with disease associated mutations are mainly buried. In contrast, leucine in general is found buried in the protein core. The leucine residues with mutations are even more buried than the leucine residues in general.

The tendencies of each amino acid to be located to a certain type of secondary structure are reflected in the mutation data so that distribution of the mutated residues in secondary structure elements, compared to distribution of the same amino acid in general, are similar (data not shown). In other words, certain secondary structure types do not occur more often among the mutated residues than the secondary structure occurs among all residues. Arginine, glycine and proline mutations are typical in loop regions even though in MAN2B1, the sites with arginine mutations are equally divided between the three secondary structure elements (Table 6). Of the five amino acids, leucine is the one most equally distributed between the different secondary structure elements (Table 6).

In addition, the distribution between the different functional areas of the protein shows features typical for each amino acid, especially for aspartic acid. Most of the aspartic acid, but also glycine residues, associated with LSDs are located in the active site area (Table 6). Instead leucine mutations do not occur prominently in the active site area, but are found on domain and dimer interfaces or locations other than the active site and interfaces (Table 6). Proline and arginine mutations

Table 6. The five most common mutation sites in the four proteins and their distribution.

Glycine	# sequence (%)	# mutations (%)	# sites (%)	Surface Mut / All	Buried Mut / All	Loop	β-strand	α-helix	Active site	Dimer interface	Domain interface	Other
GAA	72 (7.8)	26 (13.8)	20 (27.8)	3 / 44	17 / 23	16	3	1	9		7	4
GALNS	43 (8.7)	21 (15.1)	18 (42.0)	4 / 16	13 / 27	14	2	2	8	2	1	5
GLA	31 (7.8)	50 (13.0)	24 (77.4)	16 / 22	8 / 9	19	3	1	7	3	3	11
MAN2B1	65 (6.8)	7 (17.1)	6 (9.2)	1 / 33	5 / 24	3	2	1	2		2	2
all	*211*	*104*	*68*	*23*	*42*	*49*	*10*	*5*	*26*	*5*	*13*	*22*

Arginine	# sequence (%)	# mutations (%)	# sites (%)	Surface Mut / All	Buried Mut / All	Loop	β-strand	α-helix	Active site	Dimer interface	Domain interface	Other
GAA	48 (5.2)	26 (13.8)	15 (31.3)	5 / 29	10 / 14	5	7	3	3		5	7
GALNS	24 (4.8)	15 (10.9)	10 (42.0)	5 / 12	5 / 12	6		4	2	2	3	3
GLA	19 (4.8)	21 (5.4)	9 (47.3)	6 / 16	3 / 3	4	3	2	2	1	2	4
MAN2B1	65 (6.8)	7 (17.1)	5 (7.7)	1 / 42	4 / 20	1	2	2	2		2	1
all	*156*	*69*	*39*	*17*	*21*	*15*	*12*	*11*	*9*	*3*	*12*	*15*

Leucine	# sequence (%)	# mutations (%)	# sites (%)	Surface Mut / All	Buried Mut / All	Loop	β-strand	α-helix	Active site	Dimer interface	Domain interface	Other
GAA	102 (11.0)	19 (10.1)	14 (14.0)	3 / 36	11 / 60	5	3	3	1		5	8
GALNS	55 (11.0)	11 (7.9)	9 (16.0)	4 / 21	5 / 34	5	2	3	2		1	6
GLA	41 (10.3)	33 (8.5)	22 (53.7)	6 / 17	16 / 22	9	7	5	5	1	5	11
MAN2B1	101 (10.5)	6 (14.6)	6 (5.9)	2 / 42	4 / 57	3	3				4	2
all	*299*	*69*	*51*	*15*	*35*	*22*	*15*	*11*	*8*	*1*	*15*	*27*

Proline	# sequence (%)	# mutations (%)	# sites (%)	Surface Mut / All	Buried Mut / All	Loop	β-strand	α-helix	Active site	Dimer interface	Domain interface	Other
GAA	79 (8.5)	20 (10.6)	15 (36.6)	8 / 49	7 / 22	11		1	3		5	7
GALNS	41 (8.3)	12 (8.7)	9 (22.0)	2 / 21	7 / 19	5	1	3	4	1	2	2
GLA	19 (4.8)	24 (6.2)	12 (63.1)	7 / 14	5 / 5	10	1	1	4	2	2	4
MAN2B1	63 (6.5)	4 (9.8)	4 (6.3)	1 / 41	3 / 12	3		1	3			1
all	*202*	*60*	*40*	*18*	*20*	*28*	*2*	5	*14*	*3*	7	*15*

Aspartate	# sequence (%)	# mutations (%)	# sites (%)	Surface Mut / All	Buried Mut / All	Loop	β-strand	α-helix	Active site	Dimer interface	Domain interface	Other
GAA	40 (4.3)	7 (3.7)	4 (10.0)	3 / 26	4 / 11	2	1	1	4			3
GALNS	24 (4.8)	7 (5.1)	6 (25.0)	2 / 13	4 / 11	4		2	2	1	2	1
GLA	29 (7.3)	36 (9.3)	14 (48.3)	5 / 19	9 / 9	9	3	2	7	1	3	3
MAN2B1	51 (5.3)	2 (4.9)	2 (3.9)	0 / 36	2 / 13	2			2			
all	*144*	*52*	*26*	*10*	*18*	*16*	*4*	*5*	15	*2*	*5*	*7*

are distributed quite evenly between the different functional areas (Table 6).

5.2.3.3 Structural consequences of the disease-associated mutations

The structural analysis suggests that glycine, arginine, proline, aspartic acid, cysteine and histidine are over-represented among the disease associated mutations, because they have unique features that are difficult to replace by other residues.

These features may reflect the overall protein where the mutations occur. For instance aspartic acid and histidine, where disease associated mutations are both over-represented only in one protein each (GLA or MAN2B1) are typically primary active site residues (Table 6). This also explains their high frequency among disease association in GLA and MAN2B1. Seven out of the 14 aspartate mutation sites in GLA are located in the active site. Similarly, all three histidine mutations in MAN2B1 are in the active site. Disease associated glycine mutations are also often found in the active site, where the glycine is next to a primary active site residue. The replacement of glycine will almost always disturb the geometry of the active site.

Due to the small overall number of active site residues, only a few disease associated mutations locate in the active site. It has been even been proposed that LSDs should be considered as protein folding diseases (Rivera-Colón *et al*., 2012). In agreement with this idea, also my analysis finds most probable disease associated mutations in buried residues. In addition, mutations on residues in domain and dimer interfaces, which are also buried and dependent on the correct conformation, were also common (Table 6). In contrary, disease associated mutations of the typically very surface exposed residues such as lysine and glutamine are much less frequent than could be expected from their sequence abundance (Figure 10).

A bit surprisingly, leucine and other hydrophobic, typically buried residues are less common among the disease associated mutations than could be expected based on their abundance in sequence. Hydrophobic amino acids might be protected, because point mutations to a codon for hydrophobic residues often lead to its replacement by another very similar residue. The most common type of replacing amino acid in the disease associated leucine mutations is proline, which is detrimental, because proline disturbs the secondary structure elements where leucine often locates.

In contrary to leucine, arginine is not commonly replaced by a single type of amino acid, but the disease associated mutations are distributed to several kinds of replacements. This suggest that the cause of the protein failure is not on the replacing amino acid, but in the unique features of arginine, which are lost. Arginine has a large, charged side chain, which is capable of forming multiple interactions. Arginines often locate in dimer or domain interfaces or on protein surface regions with long loops where arginine connects secondary structure elements. Arginine is capable of forming salt bridges. Even though the arginine sites associated with LSD mutations often participate in salt bridges, the arginine residues with mutations are actually more often involved in hydrogen bonds than in salt bridges. With hydrogen bonds, arginine makes it important for the local structure that the arginine side chain has long reach and is capable of making several connections. The tolerated arginine mutations in MAN2B1 and GLA (p.R118C) are fully exposed. R118C participates in the salt bridge, which extends between two α-helices and is thus not so important for the conformation of the area. The MAN2B1 arginine (p.R337Q) is a non-conserved residue.

Glycine is unique in that it has no side chain, which makes it difficult to replace and replacement always increases the residue volume. The lack of side chain allows backbone conformations that are not accessible for other amino acids. Thus, typically glycine residues with disease-associated mutations are next to primary catalytic residues or in loops (Table 6).

Proline has a cyclic side chain, which limits the backbone rotation angles and is often required in protein structures where the polypeptide needs to change direction. Thus the disease-associated proline mutations occur commonly in tight turns and or loop regions, which explains why they are common on surface exposed residues. As with arginine, proline sites associated with LSDs include several different types of replacing amino acids, suggesting that the loss of proline is a more important contribution to protein failure than introduction of the replacing residue. In the buried sites with disease-associated proline mutations there was often no space for the replacing amino acid side chain. Though the surface exposed sites with proline mutations could often accommodate the side chain of the replacing amino acid, the correct conformation of the polypeptide backbone could not be achieved with the replacing amino acids. A significant proportion of the proline residues with disease-associated mutations are in the active site area (Table 6).

Mutations in cysteine residues seem inevitably disease associated. In GLA, all ten cysteine residues which participate on disulphide bridges, also contain disease associated mutations. This is in agreement with the fact that disulphide bridges are extremely important for protein structure. In GLA, many of the disulphide bridges are on loops close to active site. In GALNS, the disease associated cysteine mutation is at C79, which has the formylglycine modification and thus works as nucleophile in the active site (Bojarova and Williams, 2008).

5.2.3.4 Towards understanding therapies of lysosomal storage disorders

The effectivity of many LSD therapies depend on the type of the individual disease associated mutation. For instance, due to the immune response, enzyme replacement therapy can be less effective in patient that have no detectable level of the endogenous protein (Brooks *et al*., 2003). Enzyme enhancement therapy is often successful when the disease-associated mutation affects folding of the enzyme and thus causes premature degradation in ER or in lysosomes. Usually mutations that cause relatively mild late onset phenotype respond well to chaperone therapy. Typically for these mutations, the chemistry or volume of the resulting residue is not significantly different from the original residue. However, a significant portion of the severe classic Fabry phenotype associated mutations respond to EET (Benjamin *et al*., 2009), as long as the mutations do not severely disturb folding. Examples of mutations, which do not respond to EET , are on residues, which are responsible for the maintenance of the $(\alpha/\beta)_8$ barrel fold ,such as p.L131P (Benjamin *et al*., 2009).

Active site mutations are problematic for EET for two reasons. As the chemical chaperones are usually competitive inhibitors they bind to the same area as the substrate, but if mutation interferes with substrate binding, also chaperone may bind poorly. p.D92N/Y in GLA is an example of non-responsive substrate binding mutations (Benjamin *et al*., 2009). If the mutation directly affects the catalytic residues, substrate or chaperone binding might occur and the protein is transported to lysosomes, but it has no activity. However, also mutations on active site area may be responsive to EET. For example MAN2B1 mutation p.H72L,

which affects the metal binding residue in the active site, is responsive to Zn^{2+} treatment in patient cultured fibroblasts (Bach *et al*., 1978).

The effects of the chaperone compounds can be surprisingly long range. Missense mutations in C-terminal domain or on domain interfaces are common among GLA disease-associated mutations (Table 5). Some of these are located quite far away from the active site, where the chaperone molecule binds, yet all the tested GLA mutations in the C-terminal domain or domain interfaces responded to the chaperone treatment (Benjamin *et al*., 2009).

This rises two thoughts. Firstly the development of EET would greatly benefit from better understanding the stabilising mechanism of the chaperones as this would help to understand, what type peripheral mutations could be treated with EET. Secondly, more compounds like *N*-acetylcysteine, which stabilise GAA, but does not bind to the active site (Porto *et al*., 2012) should be developed as they might rescue mutations in the substrate binding area, which interfere with binding of the small substrate derived chaperones, but might still bind the more complex natural substrates in lysosomes.

5.3 Understanding of ligand binding is a prerequisite for chaperone design

5.3.1 Crystallisation of lysosomal α-glucosidase (unpublished results)

The three dimensional structure of GAA could be beneficial in many ways. With the available structure, we could better understand the catalytic mechanism of GAA. It would also help to directly analyse the molecular mechanism of Pompe disease mutations and be advantageous in developing the existing ERT preparations and in finding novel pharmacological chaperones for EET.

I was able to produce rhGAA crystals only after I changed the protein buffer from a phosphate to a sodium acetate buffer. Crystals appeared in a week as thick hexagonal rods, with the longest dimension being 20-50 micrometers. The best crystals diffracted X-rays to 6Å. However, the diffraction pattern was diffuse and determination of the space group or cell parameters was not possible. I made

several attempts to improve the diffraction by optimising the crystallisation conditions or by growing the crystals in presence of the GAA inhibitors deoxynojirimycin and acarbose. Unfortunately the crystal quality was not improved by optimisation, and the protein did not crystallise at all in the presence of inhibitors.

5.3.2 Systematic study of chaperone lead compounds for lysosomal α-glucosidase (Study I)

The potential of enzyme enhancement therapy (EET) for treatment of Pompe disease has been shown previously using individual inhibitors (Porto *et al*., 2009; Yoshimizu *et al*., 2008; Flanagan *et al*., 2009; Parenti *et al*., 2007). In order to study the GAA chaperone lead compounds systematically, I selected a series of 12 inhibitors (Figure 11) representing most of the available GAA competitive inhibitors. I used Myozyme as an enzyme preparation in my study because, apart

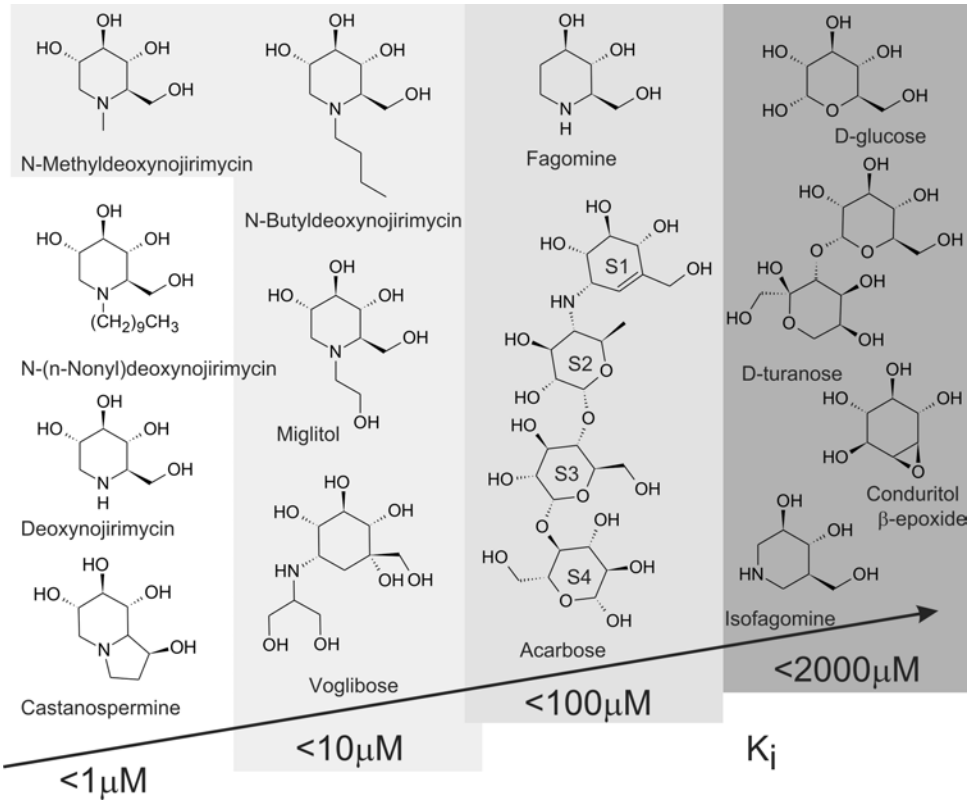


Figure 11. Inhibitors grouped by K_i values. *Background colours identify the gradient from lower to higher K_i values. We studied the binding of 13 GAA inhibitors. The binding of conduritol β-epoxide was non-competetive and it had low potential as a chaperone lead compound, it was therefore discarded from the rest of the study.*

from the cleavage of the signal sequence, it is proteolytically unmodified and thus resembles the precursor form of GAA in the ER. In addition it is used in enzyme replacement therapy and EET could be used in combination with that to stabilise the recombinant enzyme.

5.3.2.1 rhGAA stability is improved by several inhibitors

To compare the rhGAA stabilisation capacity of the inhibitors, I measured by differential scanning fluorimetry the effect of the inhibitors on rhGAA thermal unfolding at 14 different concentrations. Because an effective chaperone molecule is required to bind to GAA in the ER, rhGAA thermal stability was measured both at the acidic lysosomal pH (pH 4.3) and at neutral pH (pH 7.0 and pH 7.5) (Figure 12A). The thermal unfolding of rhGAA in the absence and presence of the inhibitors was surprisingly consistent at all pHs tested. Several inhibitors were capable of stabilising rhGAA, and all of them except isofagomine increased the unfolding temperature of rhGAA (Figure 12).

The stabilisation effect of the inhibitors appeared to be based on specific binding as the rhGAA T_m increased as a function of inhibitor concentration. D-Galactose, on the other hand, which is a stereoisomer of D-glucose but does not bind to the GAA active site, had no effect on rhGAA T_m. There were great differences in the stabilisation capacities of the different inhibitors. Castanospemine produced the highest rhGAA T_ms at all concentrations and only 2.4 µM of it was needed for the rhGAA T_m to be as high as 70°C, while a concentration of 212 mM of D-glucose was required to achieve the same effect (Figure 12C and D).

The inhibitors split into two groups, strong and weak, based on their capacity to stabilise rhGAA. These groups remained the same when inhibitors were assigned according to their inhibition constants. There was thus a clear correlation between inhibition and stabilisation. To better compare the stabilisation capacities of the individual inhibitors, I also measured the rhGAA T_m at a concentration of 10 times the K_i of each inhibitor (Table 7), thus comparing the stabilisation effect of the inhibitors under equivalent conditions.

These concentrations gave sufficient binding and the shifts in T_m were on measurable scales. Shifts in the rhGAA T_m at these concentrations showed that the stabilisation capacity of an inhibitor is related to its K_i, but the stabilisation capacity could not be derived from the K_i since the order of strength of inhibition is not the same as the order of stabilisation capacity. For instance the strongest inhibitor, deoxynojirimycin (DNJ), was the weakest stabiliser in the group of strong inhibitors. Especially interesting compounds were castanospermine and

voglibose, which showed greater stabilisation capacity than would have been expected based on their K_i,.

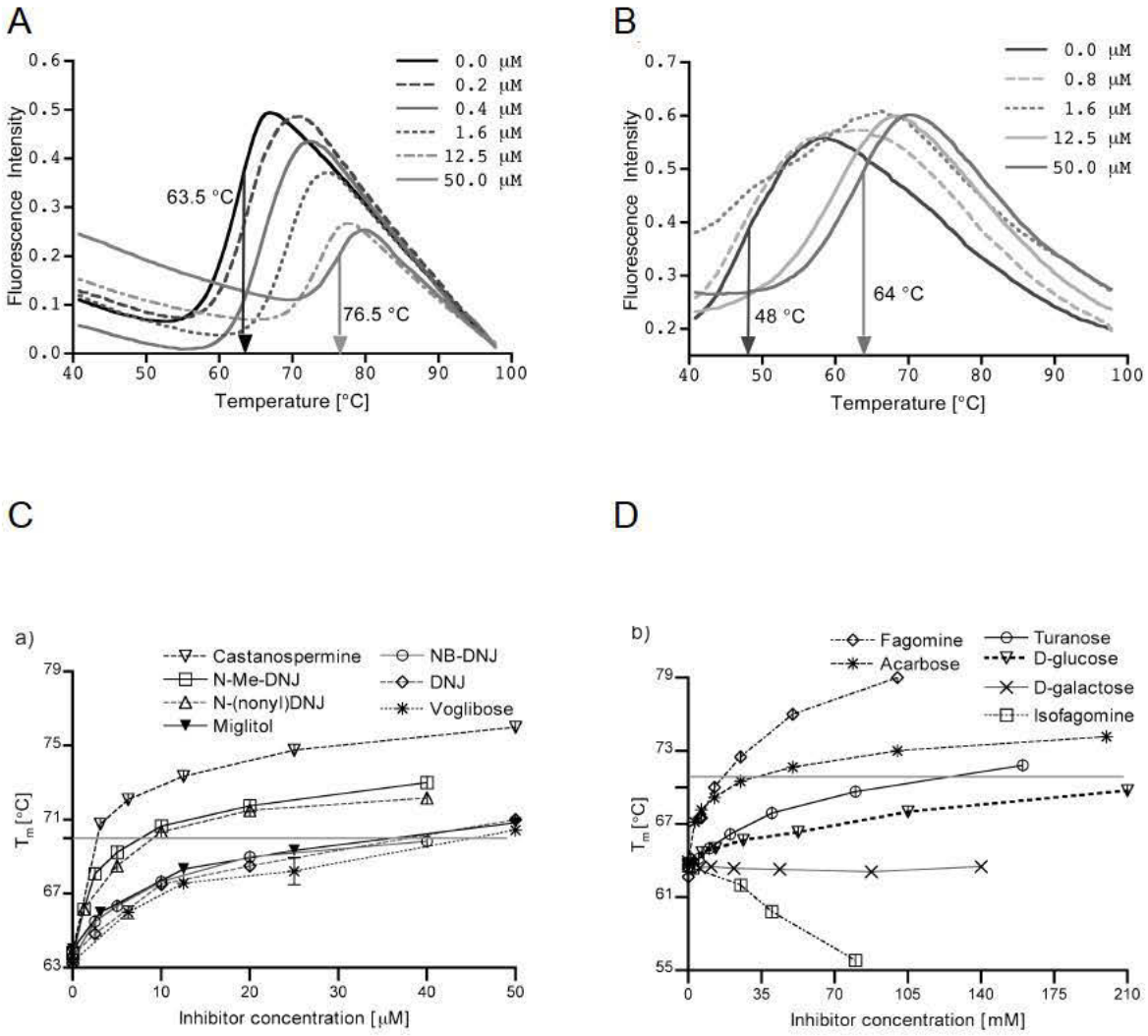


Figure 12. Thermal stabilisation of rhGAA by a selected set of inhibitors. A) and B) *Thermograms of rhGAA with castanospermine in sodium acetate buffer, pH 4.3 and in sodium phosphate buffer pH 7.5 respectively. The highest and lowest T_m of each pH are indicated with arrows. A) Castanospermine shifts the melting temperature of the rhGAA. B) phosphate buffer destabilises rhGAA and rhGAA unfolds in it at a significantly lower T_m than it does in acetate buffer. High concentrations of the inhibitor overcome this effect and restore the well-defined transition as seen in acetate buffer.* ***C) and D)*** *rhGAA T_m at pH 4.3 as a function of inhibitor concentration. The horizontal line indicates the concentration of inhibitor required for a T_m of 70°C. (C) Strong inhibitors stabilise rhGAA already at µM concentrations. D) mM concentrations of weak inhibitors are needed to see an increase in rhGAA T_m. D-galactose.*

Sodium phosphate buffer, pH 7.5, has a destabilising effect on rhGAA (Figure 12B). The unfolding of rhGAA happens over a wider temperature range than for other buffers and the Tm of rhGAA is significantly lower in phosphate buffer. I showed that the inhibitors can stabilise even unstable rhGAA because they increase the T_m and return the well defined unfolding transition as seen in other buffers (Figure 10B).

Table 7. Inhibition constants (K_i) and thermal stability of rhGAA measured in buffered solution at 10 x K_i.

				pH 4.3		pH 7.0	
Inhibitor	pI	pK_a	K_i/µM	T_m/ C	ΔT_m [a]/ C	T_m/ C	ΔT_m [a]/ C
No inhibitor	–	–	–	64.2±0.4	–	67.9±0.9	–
Conduritol b-epoxide	4.66	na	>10000	–	–	–	–
D-glucose	4.22	na	5000±500	67.3±0.3	3.10	68.4±1	0.5
Turanose	3.85	na	2000±300	67.0±0.7	2.80	69.1±1.7	1.2
Isofagomine	11.13	8.80	2000±500	63.5±0.1	-0.70	54.5±0.5	-13.4
Acarbose	9.03	7.06	210±10	67±0.3	2.80	67.0±0.3	-0.9
Fagomine	11.02	8.63	140±8	67.0±0.7	2.80	69.8±0.4	1.9
Voglibose [b]	9.99	6.77	5.8±0.2	71.8±0.3	**7.6**	71.8±0.3	**3.9**
N-butyl-deoxynojirimycin	10.67	8.49	3.1±1.0	70.4±0.3	6.20	70.7±1.1	2.8
Miglitol	10.23	7.60	3.01±0.01	70.7±0.4	6.50	68.3±0.6	0.4
N-methyl-deoxynojirimycin [b]	10.33	7.78	1.7±0.1	72.0±0.6	**7.8**	71.5±0.0	**3.6**
Castanospermine [b]	10.91	8.06	**0.68±0.08**	73.1±0.2	**8.9**	71.7±0.5	**3.8**
N-(nnonyl)-deoxynojirimycin [b]	10.33	6.37	**0.49±0.06**	69.8±1.2	5.60	70.0±0.4	2.1
Deoxynojirimycin	10.47	8.06	**0.25±0.08**	66.9±0.2	2.70	67.5±0.5	0.4

[a] ΔT_m is the difference in unfolding transition temperature for rhGAA with and without ligand ($\Delta T_m = T_m - T_0$), where T_0 is T_m at the same pH without inhibitor.
[b] The three lowest K_i values and three highest delta T_m values are in bold.

5.3.2.2 Modelling the inhibitor binding into the rhGAA active site

In order to understand the specific molecular determinants of inhibitor binding on GAA, I docked the compounds onto a homology model of GAA. The X-ray

structure of the N-terminal domain of human maltase-glucoamylase (MGA, pdb 2QMJ; Sim *et al*., 2008) was used as a template for the model. The active site is very conserved in the glycoside hydrolase family 31 and I could mainly adopt the GAA active site residue conformations and the overall binding modes of acarbose and miglitol from the MGA-ligand co-crystal structures (acarbose pdb 2QMJ and miglitol pdb 3L4W (Sim *et al*., 2010a)) (Figure 13). The binding modes of the inhibitors chemically similar to acarbose, voglibose, D-turanose and D-glucose, were modeled based on the predicted binding mode of acarbose to GAA and the inhibitors chemically similar to miglitol, DNJ and its analogs, fagomine and isofagomine, were modeled based on the miglitol binding mode.

In the model, the binding of the inhibitors to GAA is oriented by four hydroxyl groups. Acarbose is bound to the GAA active site from its acarvosin end (Figure 13A). Asp404, Asp518, His674 and Asp616 are directly hydrogen bonded to the S1 ring hydroxy groups of acarbose. In addition, Asp645 stabilises one water molecule and Asp443, Asp518 and Trp481 stabilise another water molecule that are hydrogen bonded to the hydroxyl groups of the inhibitor (Figure 13A). The hydrogen bonding of the piperidine hydroxyl groups of miglitol is similar to that of acarbose (Figure 13B) and the other inhibitors share this hydrogen bonding pattern. Isofagomine and fagomine are exceptions. They lack one of the hydroxyl groups and thus also the hydrogen bonding to Asp616. This missing hydroxyl group is the only difference between the strongest inhibitor, DNJ, and fagomine and makes the K_i of fagomine over 500 times larger than that of DNJ.

All ligands except D-glucose and D-turanose contain a secondary or tertiary amine that is predicted to be charged at neutral pH. Modeling also suggests how this affects binding to rhGAA. In acarbose, the secondary amine is between the S1 and S2 rings and Asp616 forms an ion pair with it (Figure 13A). The amine of voglibose is also stabilised by Asp616. In miglitol, the charged tertiary amine is located on the piperidine ring 2.3 Å away from the acarbose amine in the superposed structure and the charge is stabilised by Asp518 (Figure 13B). The nitrogen atom in isofagomine is part of the piperidine ring, but positioned differently to that of miglitol (Figure 11) causing it be located between Asp518 and Asp616 at hydrogen bonding distance to both, but with suboptimal bonding angles to either one. This explains why it is a much weaker inhibitor than fagomine and actually destabilises rhGAA (Table 7, Figure 12D). The inhibitors lacking the secondary or tertiary amine, D-glucose and D-turanose, are the weakest of all the inhibitors. Conversely, the strongest inhibitors have the charged nitrogen atom in the piperidine ring. Interestingly, even though voglibose has the nitrogen similarly

positioned outside the ring as acarbose, the K_i of voglibose is in the same range as that of miglitol.

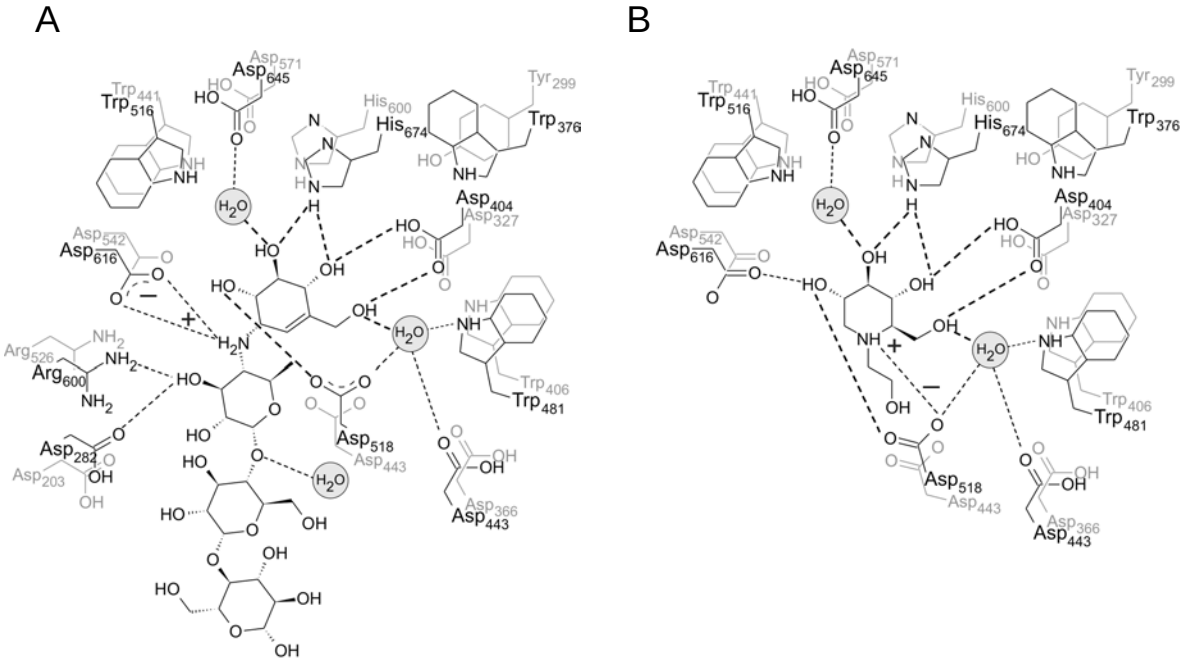


Figure 13. Superposition of the active residues of GAA and MGA. *GAA residues are presented in black and MGA residues in grey. Conserved ligand binding hydrogen bonds are emphasised with thicker lines. Neutralisation of the ligand charge is marked with + and – signs. The conformation of W376, which is not conserved in MGA, was taken from the structure of sucrose isomaltase (pdb 3LPO; Sim et al., 2010b). (A) Acarbose binding. (B) Miglitol binding.*

This work showed by thermal stability studies that rhGAA can be stabilised by several inhibitors that have varying stabilisation capacities. I demonstrated that the inhibitors also stabilise destabilised rhGAA. In addition, we showed that in a cellular environment, the most promising compounds improve the trafficking of variant GAAs to lysosomes. By combining the information gained from the modeling of the ligand binding with the inhibition studies, we suggest three factors that contribute to the tightness of the ligand binding on GAA and that should be considered when designing a chemical chaperone compound for GAA. Firstly, four well-positioned hydroxyl groups are needed to orient the ligand and make the binding specific. Secondly, as the natural substrate of GAA is a large molecule, glycogen, but very small inhibitors also bind tightly, the substrate binding cleft allows the binding of both small and large molecules. Thirdly, the

existence of a positive charge in the ligand and its location is the most important binding determinant.

In Pompe disease as in many other LSDs, EET with chemical chaperones is a promising treatment option, that is expected to help especially patients with the late onset form of the disease. The potential of EET for Pompe has been shown in *in vitro* studies (Yoshimizu *et al*., 2008), in cultured Pompe disease patient fibroblasts (Parenti *et al*., 2007) as in Pompe disease mouse model (Khanna *et al*., 2012). However, most of the studies concentrate only on a few putative chaperone compounds and useful data for developing more specific and possibly pH sensitive chaperone compounds, has been limited.

In this study, we systematically tested several putative chaperone compounds for GAA. Differential scanning fluorimetry is a good method in order to compare the stabilisation capacity of the inhibitor, as it directly gives information on thermal stabilisation of the target protein (rhGAA). Importantly, the measurements can be done also in neutral pH, which resembles the pH in the ER lumen, where *in vivo* the compounds would bind the mutated GAA. In addition, comparable rhGAA binding data with several compounds and modelling of the compounds to the GAA homology model, helped us to understand factors which determine the ligand binding strength. Combining the information should help in design of new chaperone lead compounds.

6 Conclusions and future perspectives

Lysosomes are ubiquitous cell organelles in higher eukaryotes, and they form the main degrading compartment in cells. In addition to recycling cellular components, they are involved in numerous biological functions. Lysosomal dysfunction is not only important in the individually-rare lysosomal storage disorders, but has also been recognised as a causative agent of other very common neurodegenerative diseases such as Parkinson's disease. Understanding the cellular function of lysosomes is thus crucial. This cannot be achieved without comprehending the main functional component of the lysosome, the lysosomal proteins.

The purification of bPLBD1, analysis of its post translational modifications and determination of its three dimensional structure have increased the knowledge of this protein. First of all, I showed that bPLBD1 is a soluble lysosomal hydrolase and that its glycosylation pattern contained high mannose glycans. The structure revealed that PLBD1 is an amidase. The putative substrate must have a small acyl group, but the rest of the substrate can be large and contain for example sugar moieties. This is in contrast to earlier reports, which suggest low phospholipase activity to the partially purified hPLBD1 derived from neutrophils (Xu *et al*., 2009).

Determination of the cellular function and native PLBD substrate would help in understanding if the PLBD-family is involved in an LSD of yet unknown aetiology. Identification of the PLBD substrates could be addressed in several ways. For instance, the correct molecule type could be searched for by virtual screening with molecular docking of suitable compounds with the amide bond. In my study we show that bPLBD1 has a regular thermal unfolding curve and thus, the candidate compounds could be screened for by differential scanning fluorimetry. The binging and native substrates could later be verified with enzyme activity assays.

The central undetermined aspect of M6P-dependent transport is the exact nature of the GlcNAc phosphotransferase recognition motif. I showed that the asparagine and lysine residues, which are part of the recognition site, are well conserved on the mammalian orthologous proteins, but not necessarily conserved in the fold superfamily. For bPLBD1, I was able to propose a likely candidate for the primarily phosphorylated N-glycan and the lysine residues which potentially monitor GlcNAc phosphotransferase binding. These findings need to be experimentally verified. Mutagenesis combined with cellular localisation studies

could be used to test the significance of these lysine residues for the transport of PLBDs to the lysosome.

Current treatments for LDSs are highly specific and often very expensive. Enzyme replacement therapies, when they exist, are not accessible for all patients, and also have limited effect on different types of the individual disease. Enzyme enhancement has shown great potential as a new treatment, and could possibly also be developed for Pompe disease. However, applying EET to Pompe disease would benefit from a wider range of stabilising compounds than currently known. I thus systematically studied α glucosidase inhibitors to understand the molecular determinants for effective inhibitor binding on rhGAA. The study showed that the stabilisation capacities of the different inhibitors vary greatly and do not correlate with K_i. Docking the compounds to a rhGAA homology model provided a sound structure-functional basis for development of new chemical chaperones. These studies should be continued, for instance by synthesis of related compounds as novel chaperones. A crystal structure of rhGAA with ligands, would increase the precise understanding of binding. Novel designed molecules should first be screened both *in vitro* for their ability to stabilise rhGAA, as well as by *ex vivo* to confirm the increased levels of GAA variants in chaperone treated cells. This would suggest that the chaperones could also be active *in vivo*.

Individual lysosomal storage disorders are orphan diseases and the disease associated mutations typically private, io. unique to a family. In late onset cases, with mutation carriers or in the screening of new-borns, it is difficult to know if a certain mutation is disease-associated or nonpathogenic. However, early treatment, even before the first symptoms emerge, would be beneficial to prevent serious complications. Better understanding of the general molecular mechanism by which the mutations affect lysosomal protein transport, folding or function will provide a better way of predicting the mutation severity. My analysis of the general trends of disease associated mutations will hopefully help researches and clinicians to evaluate the severity of novel mutations and also to recognise potential carriers, who are at risk to develop diseases of the elderly, such as Parkinson's.

Acknowledgements

This work was carried out for the most part in the Structural Biology and Biophysics research program at the Institute of Biotechnology, and the Division of Biochemistry in the Department of Biosciences, both in the University of Helsinki. I would like to thank the heads of the Institute during my time, Professors Mart Saarma and Tomi Mäkelä, and the head of Department, Professor Kurt Fagerstedt, for providing excellent facilities. The work was financially supported by the Academy of Finland, the Sigrid Juselius foundation, and National Doctoral Programme in Informational and Structural Biology.

I am also grateful for Professor Mark S. Johnson, the head of the National Doctoral Programme in Informational and Structural Biology, for accepting me into the programme and running this high standard programme. I would like to thank the coordinator Fredrik Karlsson for his excellent work and all past and previous members for sharing ideas and great company during the meetings.

The two people crucial for this work have been my supervisors Docent Pirkko Heikinheimo and Professor Adrian Goldman, whom both I would like to thank for excellent scientific guidance. I'm deeply grateful for Pirkko for having taken me into her group, introducing me to the scientific work and for her patience and kindness. I'm especially thankful for Adrian for adopting me to his group after Pirkko moved to Turku and providing me the excellent working conditions and support especially during writing the PLBD1 publication.

I would like to thank the members of my thesis follow-up committee. Especially Docent Eeva-Liisa Eskelinen's comments have always provided new perspectives to my work. I thank Docents Aija Kyttälä and Tiina Salminen for carefully reviewing this thesis and Professor Tapio Palva and Pekka Heino at the Department of Genetics for their help during my PhD studies. I would like thank Henri Xhaard for excellent collaboration with the GAA inhibitor studies.

In addition to having two supervisors, I have been lucky enough to belong to two research groups, and to meet so many great people. From my "original group", I would first like to thank Marika for introducing me to the subject. Elina, even though we have spent our times mostly in different cities you have greatly affected me and would like thus thank you for both the great scientific work and for our friendship. Especially, our coffee and lunch break discussions have been a joy and valuable for solving many problems concerning my studies. I would also like to thank the rest of the "Turku outstation members" especially Kisun, who I was lucky to work with during the preparation of our latest publication. My deep gratitude goes to Feri, Steffi, Kimmo, Ann-Cathrin and Johanna for their help in protein purification, thermal stability measurements and crystallisation. I want to thank Seija for her efforts with protein crystallisation and sharing her long experience in the research world and being great company for lunch. I would also like to thank Wesley, for enriching our work days by not only being the first male

member of the group, but also with his great sense of humor. I was also delighted when the post docs Andrzej and Chiara joined our group because of the scientific experience they brought with them and I'm very glad to have gotten to know these two wonderful people.

I'm also grateful for all the people who worked during the years in "my second group", Adrian's lab. Special thanks goes to Esko, Lari, Veli-Matti, Ansko, Juho, Jack, Sanjay, Vimal, Tommi, Arnab, Katja, Kornelia, and George who visited the lab twice. You all offered help already before I was a part of the group. Esko I'm thankful for you teaching me almost everything I know about X-ray crystallography and being great companion in and outside the lab. I would like to thank Tommi for having always the patience for helping me and Ansko for always finding the right words. Kornelia, I'm most grateful for our great friendship that has helped me through some days when life seemed to be too much to handle. I would also like to give my warmest thanks to Tuulia, Konstantin, Tuukka, Robert, Maike and Taru for warmly welcoming me to the group when I joined Adrian's lab after maternity leave. I would like to thank Taru especially for the last push that made it possible for me to finish and defend my thesis in this timetable. I'm also grateful for many other colleagues outside these two labs. Lotta, Annika, Nina and Jesper, it has been great to get to know all of you. Lotta I'm deeply grateful for the support in the last stretch of my thesis.

I cannot see how this work could have been possible without my other, non-scientific team, my family and friends. I'm grateful for my in-law family who has shared the ups and downs of life and for all the support they offered. The magnitude of the support from my friends has at many times left me speechless especially in recent years when life brought tragic surprises. Sharing happy moments with you has been equally important. I'm deeply grateful especially for you Ansku, Suvi, Sanna, Laura, Marja, Minna, Karo, Kati, Julia and Anu Maaret.

I'm grateful for all family members who have shown interest in my work. I owe my deepest gratitude to my parents Paula and Keijo, who always encouraged me to study, never doubted my choices and offered their support in whenever I needed it. I'm also thankful for my big brother Joni, who many times cleared my path through setting examples for me.

Finally, I would like to express my deepest gratitude and love for the two people most important in my life, my husband Aleksi and our son Konsta. You have been the biggest joy in my life and also the greatest motivation for finishing this thesis. The unshaken support of Aleksi has been needed at many times. I can't wait to begin this new period of our life and welcome the fourth member to our family.

In Helsinki, June 2014

References

Aebi, M., Bernasconi, R., Clerc, S., and Molinari, M. (2010), 'N-glycan structures: recognition and processing in the ER', *Trends Biochem Sci*, 35 (2), 74-82.

Albery, W.J. and Knowles, J.R. (1977), 'Efficiency and evolution of enzyme catalysis.', *Angew Chem Int Ed Engl*, 16 (5), 285-93.

Alkhayat, A.H., Kraemer, S.A., Leipprandt, J.R., Macek, M., Kleijer, WJ, and Friderici, KH (1998), 'Human β-mannosidase cDNA characterization and first identification of a mutation associated with human β-mannosidosis.', *Hum Mol Genet*, 7 (1), 75-83.

Aronson, N.N. Jr (1999), 'Aspartylglycosaminuria: biochemistry and molecular biology.', *Biochim Biophys Acta*, 1455 (2-3), 139-54.

Ataide, S.F., Schmitz, N., Shen, K., Ke, A., Shan, S., Doudna, J.A., and Ban, N. (2011), 'The crystal structure of the signal recognition particle in complex with its receptor.', *Science*, 331 (6019), 881-86.

Aula, P., Autio, S., Raivio, K.O., Rapola, J., Thodén, C.J., Koskela, S.L., and Yamashina, I. (1979), '"Salla disease": a new lysosomal storage disorder.', *Arch Neurol*, 36 (2), 88-94.

Ausems, M.G., Verbiest, J., Hermans, M.P., Kroos, M.A., Beemer, F.A., Wokke, J.H., Sandkuijl, L.A., Reuser, A.J., and van der Ploeg, A.T. (1999), 'Frequency of glycogen storage disease type II in The Netherlands: implications for diagnosis and genetic counselling.', *Eur J Hum Genet*, 7 (6), 713-16.

Avrahami, L., Farfara, D., Shaham-Kol, M., Vassar, R., Frenkel, D., and Eldar-Finkelman, H. (2013), 'Inhibition of glycogen synthase kinase-3 ameliorates β-amyloid pathology and restores lysosomal acidification and mammalian target of rapamycin activity in the Alzheimer disease mouse model: in vivo and in vitro studies.', *J Biol Chem*, 288 (2), 1295-306.

Baldwin, E.T., Bhat, T.N., Gulnik, S., Hosur, M.V., Sowder, R.C. 2nd, Cachau, R.E., Collins, J., Silva, A.M., and Erickson, J.W. (1993), 'Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design.', *Proc Natl Acad Sci U S A*, 90 (14), 6796-800.

Bao, M., Booth, L.J., Elmendorf, J.B., and Canfield, W.M. (1996), 'Bovine UDP-*N*-acetylglucosamine: Lysosomal-enzyme *N*-Acetylglucosamine-1-phosphotransferase I PURIFICATION AND SUBUNIT STRUCTURE', *J Biol Chem*, 271 (49), 31437-45.

Bach, G., Kohn, G., Lasch, E.E., El Massri, M., Ornoy, A., Sekeles, E., Legum, C., and Cohen, M.M. (1978), 'A new variant of mannosidosis with increased residual enzymatic activity and mild clinical manifestation', *Pediatric research*, 12 (10), 1010-15.

Barlowe, C., Orci, L., Yeung, T., Hosobuchi, M., Hamamoto, S., Salama, N., Rexach, M.F., Ravazzola, M., Amherdt, M., and Schekman, R. (1994), 'COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum', *Cell*, 77 (6), 895-907.

Baron, R., Neff, L., Louvard, D., and Courtoy, P.J. (1985), 'Cell-mediated extracellular acidification and bone resorption: evidence for a low pH in resorbing lacunae and localization of a 100-kD lysosomal membrane protein at the osteoclast ruffled border.', *J Cell Biol*, 101 (6), 2210-22.

Bause, E. (1983), 'Structural requirements of *N*-glycosylation of proteins. Studies with proline peptides as conformational probes.', *Biochem J*, 209 (2), 331-36.

Benjamin, E.R., Flanagan, J.J., Schilling, A., Chang, H.H., Agarwal, L., Katz, E., Wu, X., Pine, C., Wustman, B., and Desnick, R.J. (2009), 'The pharmacological chaperone 1-deoxygalactonojirimycin increases α-galactosidase A levels in Fabry patient cell lines', *J inher met dis*, 32 (3), 424-40.

Beratis, N.G., LaBadie, G.U., and Hirschhorn, K. (1978), 'Characterization of the molecular defect in infantile and adult acid α-glucosidase deficiency fibroblasts.', *J Clin Invest*, 62 (6), 1264-74.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000), 'The protein data bank', *Nucleic acids research*, 28 (1), 235-42.

Bernardo, K., Hurwitz, R., Zenk, T., Desnick, R.J., Ferlinz, K, Schuchman, EH, and Sandhoff, K (1995), 'Purification, characterization, and biosynthesis of human acid ceramidase.', *J Biol Chem*, 270 (19), 11098-102.

Bijvoet, A.G., Kroos, M.A., Pieper, F.R., Van der Vliet, M., De Boer, H.A., Van der Ploeg, A.T., Verbeet, M.P., and Reuser, A.J. (1998), 'Recombinant human acid α-glucosidase: high level production in mouse milk, biochemical characteristics, correction of enzyme deficiency in GSDII KO mice.', *Hum Mol Genet*, 7 (11), 1815-24.

Boelens, J.J. and R. Wynn (2013), *Stem Cell Therapy in Lysosomal Storage Diseases*, (Springer).

Bojarová, P. and Williams, S.J. (2008), 'Sulfotransferases, sulfatases and formylglycine-generating enzymes: a sulfation fascination.', *Curr Opin Chem Biol*, 12 (5), 573-81.

Bokhove, M., Yoshida, H., Hensgens, C.M., van der Laan, J.M., Sutherland, J.D., and Dijkstra, B.W. (2010), 'Structures of an isopenicillin N converting Ntn-hydrolase reveal different catalytic roles for the active site residues of precursor and mature enzyme', *Structure*, 18 (3), 301-8.

Bond, C.S., Clements, P.R., Ashby, S.J., Collyer, C.A., Harrop, S.J., Hopwood, J.J., and Guss, J.M. (1997), 'Structure of a human lysosomal sulfatase', *Structure*, 5 (2), 277-89.

Bonten, E., van der Spoel, A., Fornerod, M., Grosveld, G., and d'Azzo, A. (1996), 'Characterization of human lysosomal neuraminidase defines the molecular basis of the metabolic storage disorder sialidosis.', *Genes Dev*, 10 (24), 3156-69.

Boustany, R-M.N. (2013), 'Lysosomal storage diseases—the horizon expands', *Nature Reviews Neurology*, 9 (10), 583-98.

Boya, P. (2012), 'Lysosomal function and dysfunction: mechanism and disease', *Antioxid Redox Signal*, 17 (5), 766-74.

Brady, R.O. (2006), 'Enzyme replacement for lysosomal diseases', *Annu Rev Med*, 57 283-96.

Braulke, T. and Bonifacino, J.S. (2009), 'Sorting of lysosomal proteins', *Biochim Biophys Acta*, 1793 (4), 605-14.

Brooks, D.A., Kakavanos, R., and Hopwood, J.J. (2003), 'Significance of immune response to enzyme-replacement therapy for patients with a lysosomal storage disorder', *Trends mol med*, 9 (10), 450-53.

Bruckmann, C., Repo, H., Kuokkanen, E., Xhaard, H., and Heikinheimo, P. (2012), 'Systematic structure-activity study on potential chaperone lead compounds for acid α-glucosidase.', *ChemMedChem*, 7 (11), 1943-53.

Brömme, D., Li, Z., Barnes, M., and Mehler, E. (1999), 'Human cathepsin V functional expression, tissue distribution, electrostatic surface potential, enzymatic characterization, and chromosomal localization', *Biochemistry*, 38 (8), 2377-85.

Burda, P. and Aebi, M. (1999), 'The dolichol pathway of *N*-linked glycosylation.', *Biochim Biophys Acta*, 1426 (2), 239-57.

Canuel, M., Korkidakis, A., Konnyu, K., and Morales, C.R. (2008), 'Sortilin mediates the lysosomal targeting of cathepsins D and H.', *Biochem Biophys Res Commun*, 373 (2), 292-97.

Caramelo, J.J. and Parodi, A.J. (2008), 'Getting in and out from calnexin/calreticulin cycles.', *J Biol Chem*, 283 (16), 10221-25.

Carlsson, S.R., Roth, J., Piller, F., and Fukuda, M. (1988), 'Isolation and characterization of human lysosomal membrane glycoproteins, h-lamp-1 and h-lamp-2. Major sialoglycoproteins carrying polylactosaminoglycan.', *J Biol Chem*, 263 (35), 18911-19.

Chen, R., Jiang, X., Sun, D., Han, G., Wang, F., Ye, M., Wang, L., and Zou, H. (2009), 'Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry', *J Proteome Res*, 8 (2), 651-61.

Clark, N.E. and Garman, S.C. (2009), 'The 1.9 Å structure of human α-*N*-acetylgalactosaminidase: The molecular basis of Schindler and Kanzaki diseases.', *J Mol Biol*, 393 (2), 435-47.

Consortium, T.U. (2012), 'Reorganizing the protein space at the Universal Protein Resource (UniProt)', *Nucleic Acids Res*, 40 (Database issue), D71-75.

Coulombe, R., Grochulski, P., Sivaraman, J., Ménard, R., Mort, J.S., and Cygler, M. (1996), 'Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment', *EMBO J*, 15 (20), 5492-503.

Coutinho, M.F., Prata, M.J., and Alves, S. (2012), 'A shortcut to the lysosome: the mannose-6-phosphate-independent pathway.', *Mol Genet Metab*, 107 (3), 257-66.

Cox, TM and Cachón-González, MB (2012), 'The cellular pathology of lysosomal diseases.', *J Pathol*, 226 (2), 241-54.

Cuozzo, J.W., Tao, K., Cygler, M., Mort, J.S., and Sahagian, G.G. (1998), 'Lysine-based structure responsible for selective mannose phosphorylation of cathepsin D and cathepsin L defines a common structural motif for lysosomal enzyme targeting', *J Biol Chem*, 273 (33), 21067-76.

Dahms, N.M., Olson, L.J., and Kim, J-J. P. (2008), 'Strategies for carbohydrate recognition by the mannose 6-phosphate receptors', *Glycobiology*, 18 (9), 664-78.

de Duve, C., Pressam, B.C., Gianetto, R., Wattiaux, R., and Appelmans, F. (1955), 'Tissue fractionation studies. 6. Intracellular distribution patterns of enzymes in rat-liver tissue.', *Biochem J*, 60 (4), 604-17.

de Duve, C. and Wattiaux, R. (1966), 'Functions of lysosomes', *Annu Rev Physiol*, 28 435-92.

de Duve, C. (2005), 'The lysosome turns fifty', *Nat Cell Biol*, 7 (9), 847-49.

de Saint-Vis, B, Vincent, J, Vandenabeele, S, Vanbervliet, B, Pin, JJ, Ait-Yahia, S, Patel, S, Mattei, MG, Banchereau, J, Zurawski, S, Davoust, J, Caux, C, and Lebecque, S (1998), 'A novel lysosome-associated membrane glycoprotein, DC-LAMP, induced upon DC maturation, is transiently expressed in MHC class II compartment.', *Immunity*, 9 (3), 325-36.

Dehay, B., Martinez-Vicente, M., Caldwell, G.A., Caldwell, K.A., Yue, Z., Cookson, M.R., Klein, C., Vila, M., and Bezard, E. (2013), 'Lysosomal impairment in Parkinson's disease', *Movement Disorders*, 28 (6), 725-32.

Della Valle, M.C., Sleat, D.E., Zheng, H., Moore, D.F., Jadot, M., and Lobel, P. (2011), 'Classification of subcellular location by comparative proteomic analysis of native and density-shifted lysosomes', *Mol Cell Proteomics*, 10 (4), M110 006403.

Desnuelle, C. and Salviati, L. (2011), 'Challenges in diagnosis and treatment of late-onset Pompe disease', *Curr Opin Neurol*, 24 (5), 443-48.

Deuschl, F., Kollmann, K., von Figura, K., and Lübke, T. (2006), 'Molecular characterization of the hypothetical 66.3-kDa protein in mouse: lysosomal targeting, glycosylation, processing and tissue distribution', *FEBS Lett*, 580 (24), 5747-52.

Distler, J.J. and Jourdian, G.W. (1973), 'The purification and properties of β-galactosidase from bovine testes', *J Biol Chem*, 248 (19), 6772-80.

Dittmer, F., Pohlmann, R., and von Figura, K. (1997), 'The phosphorylation pattern of oligosaccharides in secreted procathepsin D is glycosylation site-specific and independent of the expression of mannose 6-phosphate receptors', *J Biol Chem*, 272 (2), 852-58.

Du, H., Sheriff, S., Bezerra, J., Leonova, T., and Grabowski, G.A. (1998), 'Molecular and enzymatic analyses of lysosomal acid lipase in cholesteryl ester storage disease', *Mol Genet Metab*, 64 (2), 126-34.

Durand, S., Feldhammer, M., Bonneil, E., Thibault, P., and Pshezhetsky, A.V. (2010), 'Analysis of the biogenesis of heparan sulfate acetyl-CoA: α-glucosaminide *N*-acetyltransferase provides insights into the mechanism underlying its complete deficiency in mucopolysaccharidosis IIIC', *J Biol Chem*, 285 (41), 31233-42.

Elliott, M.R. and Ravichandran, K.S. (2010), 'Clearance of apoptotic cells: implications in health and disease', *J Cell Biol*, 189 (7), 1059-70.

Eskelinen, E.L. and Saftig, P. (2009), 'Autophagy: a lysosomal degradation pathway with a central role in health and disease', *Biochim Biophys Acta*, 1793 (4), 664-73.

Fan, J.Q. (2008), 'A counterintuitive approach to treat enzyme deficiencies: use of enzyme inhibitors for restoring mutant enzyme activity', *Biol Chem*, 389 (1), 1-11.

Flanagan, J.J., Rossi, B., Tang, K., Wu, X., Mascioli, K., Donaudy, F., Tuzzi, M.R., Fontana, F., Cubellis, M.V., Porto, C., Benjamin, E., Lockhart, D.J., Valenzano, K.J., Andria, G., Parenti, G., and Do, H.V. (2009), 'The pharmacological chaperone 1-deoxynojirimycin increases the activity and lysosomal trafficking of multiple mutant forms of acid alpha-glucosidase', *Hum Mutat*, 30 (12), 1683-92.

Foltynie, T. and Kahan, J. (2013), 'Parkinson's disease: an update on pathogenesis and treatment', *J Neurol*, 260 (5), 1433-40.

Friedrichs, B., Tepel, C., Reinheckel, T., Deussing, J., von Figura, K., Herzog, V., Peters, C., Saftig, P., and Brix, K. (2003), 'Thyroid functions of mouse cathepsins B, K, and L', *J Clin Invest*, 111 (11), 1733-45.

Fujita, H., Ezaki, J., Noguchi, Y., Kono, A., Himeno, M., and Kato, K. (1991), 'Isolation and sequencing of a cDNA clone encoding 85kDa sialoglycoprotein in rat liver lysosomal membranes', *Biochem Biophys Res Commun*, 178 (2), 444-52.

Gabius, H.J., André, S., Jiménez-Barbero, J., Romero, A., and Solís, D. (2011), 'From lectin structure to functional glycomics: principles of the sugar code', *Trends Biochem Sci*, 36 (6), 298-313.

Galjart, N.J., Morreau, H., Willemsen, R., Gillemans, N., Bonten, E.J., and d'Azzo, A. (1991), 'Human lysosomal protective protein has cathepsin A-like activity distinct from its protective function', *J Biol Chem*, 266 (22), 14754-62.

Gallegos, A.M., Atshaves, B.P., Storey, S., Schoer, J., Kier, A.B., and Schroeder, F. (2002), 'Molecular and fluorescent sterol approaches to probing lysosomal membrane lipid dynamics', *Chem Phys Lipids*, 116 (1-2), 19-38.

Garman, S.C. and Garboczi, D.N. (2004), 'The molecular defect leading to Fabry disease: structure of human α-galactosidase', *J Mol Biol*, 337 (2), 319-35.

Gatt, S. (1963), 'Enzymic hydrolysis and synthesis of ceramides', *J Biol Chem*, 238 (9), PC3131-33.

Giugliani, R., Federhen, A., Rojas, M.V., Vieira, T., Artigalás, O., Pinto, L.L., Azevedo, A.C., Acosta, A., Bonfim, C., Lourenco, C.M., Kim, C.A., Horovitz, D., Bonfim, D., Norato, D., Marinho, D., Palhares, D., Santos, E.S., Ribeiro, E., Valadares, E., Guarany, F., de Lucca, G.R., Pimentel, H., de Souza, I.N., Correa, J. Sr, Fraga, J.C, Goes, J.E., Cabral, J.M., Simionato, J., Llerena, J. Jr, Jardim, L., Giuliani, L., da Silva, L.C., Santos, M.L., Moreira, M.A., Kerstenetzky, M., Ribeiro, M., Ruas, N., Barrios, P., Aranda, P., Honjo, R., Boy, R., Costa, R., Souza, C., Alcantara, F.F., Avilla, S.G., Fagondes, S., and Martins, A.M. (2010), 'Mucopolysaccharidosis I, II, and VI: Brief review and guidelines for treatment', *Genet Mol Biol*, 33 (4), 589-604.

Glamuzina, E., Fettes, E., Bainbridge, K., Crook, V., Finnegan, N., Abulhoul, L., and Vellodi, A. (2011), 'Treatment of mucopolysaccharidosis type II (Hunter syndrome) with idursulfase: the relevance of clinical trial end points', *J. Inherit. Metab. Dis.*, 34 (3), 749-54.

Glerup, S., Lume, M., Olsen, D., Nyengaard, J.R., Vaegter, C.B., Gustafsen, C., Christensen, E.I., Kjolby, M., Hay-Schmidt, A., Bender, D., Madsen, P., Saarma, M., Nykjaer, A., and Petersen, C.M. (2013), 'SorLA controls neurotrophic activity by sorting of GDNF and its receptors GFRα1 and RET', *Cell Rep*, 3 (1), 186-99.

Gromiha, M.M., Pathak, M.C., Saraboji, K., Ortlund, E.A., and Gaucher, E.A. (2013), 'Hydrophobic environment is a key factor for the stability of thermophilic proteins', *Proteins*, 81 (4), 715-21.

Guce, A.I., Clark, N.E., Salgado, E.N., Ivanen, D.R., Kulminskaya, A.A., Brumer, H. 3rd, and Garman, S.C. (2010), 'Catalytic mechanism of human α--galactosidase', *J Biol Chem*, 285 (6), 3625-32.

Hamer, I., Van Beersel, G., Arnould, T., and Jadot, M. (2012), 'Lipids and lysosomes', *Curr Drug Metab*, 13 (10), 1371-87.

Hammond, C, Braakman, I, and Helenius, A (1994), 'Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control.', *Proc Natl Acad Sci U S A*, 91 (3), 913-17.

Hansen, G., Berg, T., Riise Stensland, H.M., Heikinheimo, P., Klenow, H., Evjen, G., Nilssen, O., and Tollersrud, O.K. (2004), 'Intracellular transport of human lysosomal α-mannosidase and α-mannosidosis-related mutants', *Biochem J*, 381 (Pt 2), 537-46.

Hanson, S.R., Culyba, E.K., Hsu, T.L., Wong, C.H, Kelly, J.W., and Powers, E.T. (2009), 'The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability', *Proc Natl Acad Sci U S A*, 106 (9), 3131-36.

Harada, Y., Li, H., Li, H., and Lennarz, W.J. (2009), 'Oligosaccharyltransferase directly binds to ribosome at a location near the translocon-binding site', *Proc Natl Acad Sci U S A*, 106 (17), 6945-49.

Haskins, M.E., Jezyk, P.F., and Patterson, D.F. (1979), 'Mucopolysaccharide storage disease in three families of cats with arylsulfatase B deficiency: leukocyte studies and carrier identification', *Pediatr Res*, 13 (11), 1203-10.

Hatem, C.L., Gough, N.R., and Fambrough, D.M. (1995), 'Multiple mRNAs encode the avian lysosomal membrane protein LAMP-2, resulting in alternative transmembrane and cytoplasmic domains', *J Cell Sci*, 108 (Pt 5), 2093-100.

Heikinheimo, P., Helland, R., Leiros, H.K., Leiros, I., Karlsen, S., Evjen, G., Ravelli, R., Schoehn, G., Ruigrok, R., Tollersrud, O.K., McSweeney, S., and Hough, E. (2003), 'The structure of bovine lysosomal α-mannosidase suggests a novel mechanism for low-pH activation', *J Mol Biol*, 327 (3), 631-44.

Helenius, A. and Aebi, M. (2004), 'Roles of N-linked glycans in the endoplasmic reticulum', *Annu Rev Biochem*, 73 1019-49.

Henrissat, B. and Davies, G. (1997), 'Structural and sequence-based classification of glycoside hydrolases', *Curr Opin Struct Biol*, 7 (5), 637-44.

Hermans, M.M., Kroos, M.A., van Beeumen, J., Oostra, B.A., and Reuser, A.J. (1991), 'Human lysosomal α-glucosidase. Characterization of the catalytic site.', *J Biol Chem*, 266 (21), 13507-12.

Hers, H.G. (1963), 'α-Glucosidase deficiency in generalized glycogenstorage disease (Pompe's disease)', *Biochem J*, 86 11-16.

Hickman, S. and Neufeld, E.F. (1972), 'A hypothesis for I-cell disease: defective hydrolases that do not enter lysosomes', *Biochem Biophys Res Commun*, 49 (4), 992-99.

Hiraoka, M., Abe, A., and Shayman, J.A. (2002), 'Cloning and characterization of a lysosomal phospholipase A2, 1-*O*-acylceramide synthase', *J Biol Chem*, 277 (12), 10090-99.

Hobbs, J.R., Hugh-Jones, K., Barrett, A.J., Byrom, N., Chambers, D., Henry, K., James, D.C., Lucas, C.F., Rogers, T.R., Benson, P.F., Tansley, L.R., Patrick, A.D., Mossman, J., and Young, E.P. (1981), 'Reversal of clinical features of Hurler's disease and biochemical improvement after treatment by bone-marrow transplantation', *Lancet*, 2 (8249), 709-12.

Hoogeveen, A.T., Reuser, A.J., Kroos, M., and Galjaard, H. (1986), 'GM1-gangliosidosis. Defective recognition site on β-galactosidase precursor', *J Biol Chem*, 261 (13), 5702-4.

Hopwood, J.J. (2012), 'Genetics of Lysosomal Storage Disorders and Counselling', in Mehta, A. and Winchester, B (eds.), (Lysosomal Storage Disorders: A Practical Guide, Wiley Online Library), 29-36.

Howell, D.P., Krieser, R.J., Eastman, A., and Barry, M.A. (2003), 'Deoxyribonuclease II is a lysosomal barrier to transfection', *Mol Ther*, 8 (6), 957-63.

Huber, R., Langworthy, T.A., König, H., Thomm, M., Woese, C.R., Sleytr, U.B., and Stetter, K.O. (1986), '*Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C', *Arch Microbiol*, 144 (4), 324-33.

Hruban, Z., Spargo, B., Swift, H., Wissler, R.W., and Kleinfeld, R.G. (1963), 'Focal cytoplasmic degradation', *Am J Pathol*, 42 657-83.

Irvine, R.F., Hemington, N., and Dawson, R.M. (1978), 'The hydrolysis of phosphatidylinositol by lysosomal enzymes of rat liver and brain', *Biochem J*, 176 (2), 475-84.

Jansen, G., Määttänen, P., Denisov, A.Y., Scarffe, L., Schade, B., Balghi, H., Dejgaard, K., Chen, L.Y., Muller, W.J., Gehring, K., and Thomas, D.Y. (2012), 'An interaction map of endoplasmic reticulum chaperones and foldases', *Mol Cell Proteomics*, 11 (9), 710-23.

Jensen, A.G., Chemali, M., Chapel, A., Kieffer-Jaquinod, S., Jadot, M., Garin, J., and Journet, A. (2007), 'Biochemical characterization and lysosomal localization of the mannose-6-phosphate protein p76 (hypothetical protein LOC196463)', *Biochem J*, 402 (3), 449-58.

Jiang, S., Heller, B., Tagliabracci, V.S., Zhai, L., Irimia, J.M., DePaoli-Roach, A.A., Wells, C.D., Skurat, A.V., and Roach, P.J. (2010), 'Starch binding domain-containing protein 1/ genethonin 1 is a novel participant in glycogen metabolism.', *J Biol Chem*, 285 (45), 34960-71.

Johnson, S.W. and Alhadeff, J.A. (1991), 'Mammalian α-L-fucosidases', *Comp Biochem Physiol B*, 99 (3), 479-88.

Journet, A., Chapel, A., Kieffer, S., Roux, F., and Garin, J. (2002), 'Proteomic analysis of human lysosomes: application to monocytic and breast cancer cells', *Proteomics*, 2 (8), 1026-40.

Kallunki, T., Olsen, O.D., and Jäättelä, M. (2013), 'Cancer-associated lysosomal changes: friends or foes?', *Oncogene*, 32 (16), 1995-2004.

Kang, C., Riazuddin, S., Mundorff, J., Krasnewich, D., Friedman, P., Mullikin, J.C., and Drayna, D. (2010), 'Mutations in the lysosomal enzyme-targeting pathway and persistent stuttering.', *N Engl J Med*, 362 (8), 677-85.

Khanna, R., Flanagan, J.J., Feng, J., Soska, R., Frascella, M., Pellegrino, L.J., Lun, Y., Guillen,

D., Lockhart, D.J., and Valenzano, K.J. (2012), 'The pharmacological chaperone AT2220 increases recombinant human acid α-glucosidase uptake and glycogen reduction in a mouse model of Pompe disease', *PloS one*, 7 (7), e40776.

Kim, I., Rodriguez-Enriquez, S., and Lemasters, J.J. (2007), 'Selective degradation of mitochondria by mitophagy', *Arch Biochem Biophys*, 462 (2), 245-53.

Kishnani, P.S., Corzo, D., Nicolino, M., Byrne, B., Mandel, H., Hwu, W.L., Leslie, N., Levine, J., Spencer, C., McDonald, M., Li, J., Dumontier, J., Halberthal, M., Chien, Y.H., Hopkin, R., Vijayaraghavan, S., Gruskin, D., Bartholomew, D., van der Ploeg, A., Clancy, J.P., Parini, R., Morin, G., Beck, M., De la Gastine, G.S., Jokic, M., Thurberg, B., Richards, S., Bali, D., Davison, M., Worden, M.A., Chen, Y.T., and Wraith, J.E. (2007), 'Recombinant human acid α-glucosidase: major clinical benefits in infantile-onset Pompe disease', *Neurology*, 68 (2), 99-109.

Kollmann, K., Mutenda, K.E., Balleininger, M., Eckermann, E., von Figura, K., Schmidt, B., and Lübke, T. (2005), 'Identification of novel lysosomal matrix proteins by proteome analysis', *PROTEOMICS*, 5 (15), 3966-78.

Korneluk, R.G., Mahuran, D.J., Neote, K., Klavins, M.H., O'Dowd, B.F., Tropak, M., Willard, H.F., Anderson, M.J., Lowden, J.A., and Gravel, R.A. (1986), 'Isolation of cDNA clones coding for the α-subunit of human β-hexosaminidase. Extensive homology between the α- and β-subunits and studies on Tay-Sachs disease', *J Biol Chem*, 261 (18), 8407-13.

Kornfeld, R., Bao, M., Brewer, K., Noll, C., and Canfield, W. (1999), 'Molecular cloning and functional expression of two splice forms of human *N*-acetylglucosamine-1-phosphodiester α-*N*-acetylglucosaminidase', *J Biol Chem*, 274 (46), 32778-85.

Korolchuk, V.I., Menzies, F.M., and Rubinsztein, D.C. (2010), 'Mechanisms of cross-talk between the ubiquitin-proteasome and autophagy-lysosome systems', *FEBS Lett*, 584 (7), 1393-98.

Korpela, M.P., Paetau, A., Löfberg, M.I., Timonen, M.H., Lamminen, A.E., and Kiuru-Enari, S.MK. (2009), 'A novel mutation of the GAA gene in a Finnish late-onset pompe disease patient: Clinical phenotype and follow-up with enzyme replacement therapy', *Muscle & Nerve*, 40 (1), 143-48.

Kowarik, M., Küng, S., Martoglio, B., and Helenius, A. (2002), 'Protein folding during cotranslational translocation in the endoplasmic reticulum', *Mol Cell*, 10 (4), 769-78.

Kroos, M., Pomponio, R.J., van Vliet, L., Palmer, R.E., Phipps, M., Van der Helm, R., Halley, D., and Reuser, A. (2008), 'Update of the Pompe disease mutation database with 107 sequence variants and a format for severity rating.', *Hum Mutat*, 29 (6), E13-26.

Kroos, M., Hoogeveen-Westerveld, M., van der Ploeg, A., and Reuser, A.J.J. (2012), 'The genotype–phenotype correlation in Pompe disease', Am J Med Genet Part C: Seminars in Medical Genetics 160(1) (1), 59-68.

Kumar, S., Tsai, C-J., and Nussinov, R. (2000), 'Factors enhancing protein thermostability', *Protein Engineering*, 13 (3), 179-91.

Kume, S., Yamahara, K., Yasuda, M., Maegawa, H., and Koya, D. (2014), 'Autophagy: emerging therapeutic target for diabetic nephropathy', *Semin Nephrol*, 34 (1), 9-16.

Kuokkanen, E., Riise Stensland, H.M., Smith, W., Kjeldsen Buvang, E., Van Nguyen, L., Nilssen, O., and Heikinheimo, P. (2011), 'Molecular and cellular characterization of novel α-mannosidosis mutations.', *Hum Mol Genet*, 20 (13), 2651-61.

Labrousse, P., Chien, Y.H., Pomponio, R.J., Keutzer, J., Lee, N.C., Akmaev, V.R., Scholl, T., and Hwu, W.L. (2010), 'Genetic heterozygosity and pseudodeficiency in the Pompe disease newborn screening pilot program', *Mol Genet Metab*, 99 (4), 379-83.

Lakomek, K., Dickmanns, A., Mueller, U., Kollmann, K., Deuschl, F., Berndt, A., Lübke, T., and Ficner, R. (2009), 'De novo sulfur SAD phasing of the lysosomal 66.3 kDa protein from mouse', *Acta Crystallogr D Biol Crystallogr*, 65 (Pt 3), 220-28.

Lambert, N. and Freedman, R.B. (1985), 'The latency of rat liver microsomal protein disulphide-isomerase', *Biochem J*, 228 (3), 635-45.

Laurent, T.C. and Fraser, J.R. (1992), 'Hyaluronan', *FASEB J*, 6 (7), 2397-404.

Lederkremer, G.Z. and Glickman, M.H. (2005), 'A window of opportunity: timing protein degradation by trimming of sugars and ubiquitins', *Trends Biochem Sci*, 30 (6), 297-303.

Lederkremer, G.Z. (2009), 'Glycoprotein folding, quality control and ER-associated degradation', *Curr Opin Struct Biol*, 19 (5), 515-23.

LeDonne Jr, N.C., Fairley, J.L., and Sweeley, C.C. (1983), 'Biosynthesis of α-galactosidase A in cultured Chang liver cells', *Archi Biochem Biophys*, 224 (1), 186-95.

Lee, K.S., Patton, J.L., Fido, M., Hines, L.K., Kohlwein, S.D., Paltauf, F., Henry, S.A., and Levin, D.E. (1994), 'The Saccharomyces cerevisiae PLB1 gene encodes a protein required for lysophospholipase and phospholipase B activity', *J Biol Chem*, 269 (31), 19725-30.

Lee, W.L., Harrison, R.E., and Grinstein, S. (2003), 'Phagocytosis by neutrophils', *Microbes Infect*, 5 (14), 1299-306.

Lefrancois, S., Zeng, J., Hassan, A.J., Canuel, M., and Morales, C.R. (2003), 'The lysosomal trafficking of sphingolipid activator proteins (SAPs) is mediated by sortilin.', *EMBO J*, 22 (24), 6430-37.

Lemansky, P., Bishop, D.F., Desnick, R.J., Hasilik, A., and von Figura, K. (1987), 'Synthesis and processing of α-galactosidase A in human fibroblasts. Evidence for different mutations in Fabry disease.', *J Biol Chem*, 262 (5), 2062-65.

Lieberman, R.L., D'aquino, J.A., Ringe, D., and Petsko, G.A. (2009), 'Effects of pH and iminosugar pharmacological chaperones on lysosomal glycosidase structure and stability', *Biochemistry*, 48 (22), 4816-27.

Lin, L., Sohar, I., Lackland, H., and Lobel, P. (2001), 'The human CLN2 protein/tripeptidyl-peptidase I is a serine protease that autoactivates at acidic pH.', *J Biol Chem*, 276 (3), 2249-55.

Lourenço, C.M. and Giugliani, R. (2014), 'Evaluation of galsulfase for the treatment of mucopolysaccharidosis VI (Maroteaux-Lamy syndrome)', *Expert Opin Orphan Drugs*, 2 (4), 407-17.

Lukatela, G., Krauss, N., Theis, K., Selmer, T., Gieselmann, V., von Figura, K., and Saenger, W. (1998), 'Crystal structure of human arylsulfatase A: the aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis', *Biochemistry*, 37 (11), 3654-64.

Maier, T., Strater, N., Schuette, C.G., Klingenstein, R., Sandhoff, K., and Saenger, W. (2003), 'The X-ray crystal structure of human β-hexosaminidase B provides new insights into Sandhoff disease.', *J Mol Biol*, 328 (3), 669-81.

Makrypidi, G., Damme, M., Müller-Loennies, S., Trusch, M., Schmidt, B., Schlüter, H., Heeren, J., Lübke, T., Saftig, P., and Braulke, T. (2012), 'Mannose 6 dephosphorylation of lysosomal proteins mediated by acid phosphatases Acp2 and Acp5.', *Mol Cell Biol*, 32 (4), 774-82.

Mancini, G.M., Beerens, C.E., Aula, P.P., and Verheijen, F.W. (1991), 'Sialic acid storage diseases. A multiple lysosomal transport defect for acidic monosaccharides.', *J Clin Invest*, 87 (4), 1329-35.

Marsden, D. and Levy, H. (2010), 'Newborn screening of lysosomal storage disorders', *Clinical chemistry*, 56 (7), 1071-79.

Matsumura, M., Signor, G., and Matthews, B.W. (1989), 'Substantial increase of protein stability by multiple disulphide bonds', *Nature*, 342 (6247), 291-93.

Meikle, P.J., Hopwood, J.J., Clague, A.E., and Carey, W.F. (1999), 'Prevalence of lysosomal storage disorders', *JAMA*, 281 (3), 249-54.

Meléndez, R., Meléndez-Hevia, E., and Cascante, M. (1997), 'How did glycogen structure evolve to satisfy the requirement for rapid mobilization of glucose? A problem of physical constraints in structure building', *J Mol Evol*, 45 (4), 446-55.

Mellman, I., Fuchs, R., and Helenius, A. (1986), 'Acidification of the endocytic and exocytic pathways.', *Annu Rev Biochem*, 55 663-700.

Micsenyi, M.C. and Walkley, S.U. (2012), 'The Lysosomal System: Physiology and Pathology', (Lysosomal Storage Disorders: A Practical Guide, John Wiley & Sons), 1.

Moreland, R.J., Jin, X., Zhang, X.K., Decker, R.W., Albee, K.L., Lee, K.L., Cauthron, R.D., Brewer, K., Edmunds, T., and Canfield, W.M. (2005), 'Lysosomal acid α-glucosidase consists of four different peptides processed from a single chain precursor', *J Biol Chem*, 280 (8), 6780-91.

Moremen, K.W. (2002), 'Golgi α-mannosidase II deficiency in vertebrate systems: implications for asparagine-linked oligosaccharide processing in mammals.', *Biochim Biophys Acta*, 1573 (3), 225-35.

Moremen, K.W. and Molinari, M. (2006), '*N*-linked glycan recognition and processing: the molecular basis of endoplasmic reticulum quality control.', *Curr Opin Struct Biol*, 16 (5), 592-99.

Morgan, C.P., Insall, R., Haynes, L., and Cockcroft, S. (2004), 'Identification of phospholipase B from *Dictyostelium discoideum* reveals a new lipase family present in mammals, flies and nematodes, but not yeast', *Biochem J*, 382 (Pt 2), 441-49.

Morrone, A, Tylee, K.L., Al-Sayed, M., Brusius-Facchin, A.C., Caciotti, A., Church, H.J., Coll, M.J., Davidson, K., Fietz, M.J., Gort, L., Hegde, M., Kubaski, F., Lacerda, L., Laranjeira, F., Leistner-Segal, S., Mooney, S., Pajares, S., Pollard, L., Ribeiro, I., Wang, R.Y., and Miller, N. (2014), 'Molecular testing of 163 patients with Morquio A (Mucopolysaccharidosis IVA) identifies 39 novel *GALNS* mutations', *Mol Gen and Met*,

Müller, S., Dennemärker, J., and Reinheckel, T. (2012), 'Specific functions of lysosomal proteases in endocytic and autophagic pathways', *Biochim Biophys Acta*, 1824 (1), 34-43.

Möbius, W., van Donselaar, E., Ohno-Iwashita, Y., Shimada, Y., Heijnen, H.F., Slot, J.W., and Geuze, H.J. (2003), 'Recycling compartments and the internal vesicles of multivesicular bodies harbor most of the cholesterol found in the endocytic pathway.', *Traffic*, 4 (4), 222-31.

Nagano, S., Yamada, T., Shinnoh, N., Furuya, H., Taniwaki, T., and Kira, J. (1998), 'Expression and processing of recombinant human galactosylceramidase', *Clin Chim Acta*, 276 (1), 53-61.

Nalls, M.A., Duran, R., Lopez, G., Kurzawa-Akanbi, M., McKeith, I.G., Chinnery, P.F., Morris, C.M., Theuns, J., Crosiers, D., Cras, P., Engelborghs, S., De Deyn, P.P., Van Broeckhoven, C., Mann, D.M., Snowden, J., Pickering-Brown, S., Halliwell, N., Davidson, Y., Gibbons, L., Harris, J., Sheerin, U.M., Bras, J., Hardy, J., Clark, L., Marder, K., Honig, L.S., Berg, D., Maetzler, W., Brockmann, K., Gasser, T., Novellino, F., Quattrone, A., Annesi, G., De Marco, E.V., Rogaeva, E., Masellis, M., Black, S.E., Bilbao, J.M., Foroud, T., Ghetti, B., Nichols, W.C., Pankratz, N., Halliday, G., Lesage, S., Klebe, S., Durr, A., Duyckaerts, C., Brice, A., Giasson, B.I., Trojanowski, J.Q., Hurtig, H.I., Tayebi, N., Landazabal, C., Knight, M.A., Keller, M., Singleton, A.B., Wolfsberg, T.G., and Sidransky, E. (2013), 'A multicenter study of glucocerebrosidase mutations in dementia with Lewy bodies', *JAMA Neurol*, 70 (6), 727-35.

Nichols, W.C., Pankratz, N., Marek, D.K., Pauciulo, M.W., Elsaesser, V.E., Halter, C.A., Rudolph, A., Wojcieszek, J., Pfeiffer, R.F., and Foroud, T. (2009), 'Mutations in *GBA* are associated with familial Parkinson disease susceptibility and age at onset', *Neurology*, 72 (4), 310-16.

Nykjaer, A., Lee, R., Teng, K.K., Jansen, P., Madsen, P., Nielsen, M.S., Jacobsen, C., Kliemannel, M., Schwarz, E., Willnow, T.E., Hempstead, B.L., and Petersen, C.M. (2004), 'Sortilin is essential for proNGF-induced neuronal cell death', *Nature*, 427 (6977), 843-48.

Ohashi, T (2012), 'Enzyme replacement therapy for lysosomal storage diseases.', *Pediatric endocrinology reviews: PER*, 10 26-34.

Ohkouchi, S., Shibata, M., Sasaki, M., Koike, M., Safig, P., Peters, C., Nagata, S., and Uchiyama, Y. (2013), 'Biogenesis and proteolytic processing of lysosomal DNase II', *PLoS One*, 8 (3), e59148.

Ohto, U., Usui, K., Ochi, T., Yuki, K., Satow, Y., and Shimizu, T. (2012), 'Crystal structure of human β-galactosidase: structural basis of Gm1 gangliosidosis and morquio B diseases.', *J Biol Chem*, 287 (3), 1801-12.

Oinonen, C., Tikkanen, R., Rouvinen, J., and Peltonen, L. (1995), 'Three-dimensional structure of human lysosomal aspartylglucosaminidase', *Nat Struct Biol*, 2 (12), 1102-8.

Oinonen, C. and Rouvinen, J. (2000), 'Structural comparison of Ntn-hydrolases', *Protein Sci*, 9 (12), 2329-37.

Olson, L.J., Zhang, J., Lee, Y.C., Dahms, N.M., and Kim, J-J.P. (1999), 'Structural basis for recognition of phosphorylated high mannose oligosaccharides by the cation-dependent mannose 6-phosphate receptor', *J Biol Chem*, 274 (42), 29889-96.

Orlikowski, D., Pellegrini, N., Prigent, H., Laforet, P., Carlier, R., Carlier, P., Eymard, B., Lofaso, F., and Annane, D. (2011), 'Recombinant human acid alpha-glucosidase (rhGAA) in adult patients with severe respiratory failure due to Pompe disease.', *Neuromuscul Disord*, 21 (7), 477-82.

Osborne, A.R., Rapoport, T.A., and van den Berg, B. (2005), 'Protein translocation by the Sec61/SecY channel', *Annu Rev Cell Dev Biol*, 21 529-50.

Osellame, L.D., Rahim, A.A., Hargreaves, I.P., Gegg, M.E., Richard-Londt, A., Brandner, S., Waddington, S.N., Schapira, A.H., and Duchen, M.R. (2013), 'Mitochondria and quality control defects in a mouse model of Gaucher disease - links to Parkinson's disease', *Cell Metab*, 17 (6), 941-53.

Parenti, G., Zuppaldi, A., Gabriela Pittis, M., Rosaria Tuzzi, M., Annunziata, I., Meroni, G., Porto, C., Donaudy, F., Rossi, B., Rossi, M., Filocamo, M., Donati, A., Bembi, B., Ballabio, A., and Andria, G. (2007), 'Pharmacological enhancement of mutated α-glucosidase activity in fibroblasts from patients with Pompe disease', *Mol Ther*, 15 (3), 508-14.

Parenti, Giancarlo (2009), 'Treating lysosomal storage diseases with pharmacological chaperones: from concept to clinics', *EMBO molecular medicine*, 1 (5), 268-79.

Parenti, G., Pignata, C., Vajro, P., and Salerno, M. (2013), 'New strategies for the treatment of lysosomal storage diseases', *Int J Mol Med*, 31 (1), 11-20.

Peres, G.B., Juliano, M.A., Simoes, M.J., and Michelacci, Y.M. (2013), 'Lysosomal enzymes are decreased in the kidney of diabetic rats', *Biochim Biophys Acta*, 1832 (1), 85-95.

Petersen, C.M., Nielsen, M.S., Nykjaer, A., Jacobsen, L., Tommerup, N., Rasmussen, H.H., Røigaard, H., Gliemann, J., Madsen, P., and Moestrup, S.K. (1997), 'Molecular identification of a novel candidate sorting receptor purified from human brain by receptor-associated protein affinity chromatography', *J Biol Chem*, 272 (6), 3599-605.

Pinto, R., Caseiro, C., Lemos, M., Lopes, L., Fontes, A., Ribeiro, H., Pinto, E., Silva, E., Rocha, S., Marcao, A., Ribeiro, I., Lacerda, L., Ribeiro, G., Amaral, O., and Sa Miranda, M.C. (2004), 'Prevalence of lysosomal storage diseases in Portugal', *Eur J Hum Genet*, 12 (2), 87-92.

Platt, F.M., Boland, B., and van der Spoel, A.C. (2012), 'The cell biology of disease: lysosomal storage disorders: the cellular impact of lysosomal dysfunction', *J Cell Biol*, 199 (5), 723-34.

Pohlmann, R., Waheed, A., Hasilik, A., and Von Figura, K. (1982), 'Synthesis of phosphorylated recognition marker in lysosomal enzymes is located in the *cis* part of Golgi apparatus', *J Biol Chem*, 257 (10), 5323-25.

Pohlmann, R., Boeker, M.W., and von Figura, K. (1995), 'The two mannose 6-phosphate receptors transport distinct complements of lysosomal proteins', *J Biol Chem*, 270 (45), 27311-18.

Porto, C., Cardone, M., Fontana, F., Rossi, B., Tuzzi, M.R., Tarallo, A., Barone, M.V., Andria, G., and Parenti, G. (2009), 'The pharmacological chaperone N-butyldeoxynojirimycin enhances enzyme replacement therapy in Pompe disease fibroblasts', *Mol Ther*, 17 (6), 964-71.

Porto, C., Ferrara, M.C., Meli, M., Acampora, E., Avolio, V., Rosa, M., Cobucci-Ponzano, B., Colombo, G., Moracci, M., Andria, G., and Parenti, G. (2012), 'Pharmacological enhancement of α-glucosidase by the allosteric chaperone *N*-acetylcysteine', *Mol Ther*, 20 (12), 2201-11.

Raas-Rothschild, A., Cormier-Daire, V., Bao, M., Genin, E., Salomon, R., Brewer, K., Zeigler, M., Mandel, H., Toth, S., Roe, B., Munnich, A., and Canfield, W.M. (2000), 'Molecular basis of variant pseudo-hurler polydystrophy (mucolipidosis IIIC)', *J Clin Invest*, 105 (5), 673-81.

Raben, N., Hill, V., Shea, L., Takikita, S., Baum, R., Mizushima, N., Ralston, E., and Plotz, P. (2008), 'Suppression of autophagy in skeletal muscle uncovers the accumulation of ubiquitinated proteins and their potential role in muscle damage in Pompe disease', *Hum Mol Genet*, 17 (24), 3897-908.

Rawlings, N.D., Waller, M., Barrett, A.J., and Bateman, A. (2013), 'MEROPS: the database of proteolytic enzymes, their substrates and inhibitors', *Nucleic Acids Res*,

Reczek, D., Schwake, M., Schröder, J., Hughes, H., Blanz, J., Jin, X., Brondyk, W., Van Patten, S., Edmunds, T., and Saftig, P. (2007), 'LIMP-2 is a receptor for lysosomal mannose-6-phosphate-independent targeting of β-glucocerebrosidase', *Cell*, 131 (4), 770-83.

Reitman, M.L. and Kornfeld, S. (1981), 'Lysosomal enzyme targeting. *N*-Acetylglucosaminylphosphotransferase selectively phosphorylates native lysosomal enzymes', *J Biol Chem*, 256 (23), 11977-80.

Repo, H., Oeemig, J.S., Djupsjobacka, J., Iwai, H., and Heikinheimo, P. (2012), 'A monomeric TIM-barrel structure from *Pyrococcus furiosus* is optimized for extreme temperatures', *Acta Crystallogr D Biol Crystallogr*, 68 (Pt 11), 1479-87.

Richter, K., Haslbeck, M., and Buchner, J. (2010), 'The heat shock response: life on the verge of death', *Mol Cell*, 40 (2), 253-66.

Rijnboutt, S., Aerts, H.M., Geuze, H.J., Tager, J.M., and Strous, G.J. (1991), 'Mannose 6-phosphate-independent membrane association of cathepsin D, glucocerebrosidase, and sphingolipid-activating protein in HepG2 cells.', *J Biol Chem*, 266 (8), 4862-68.

Rivera-Colón, Y., Schutsky, E.K., Kita, A.Z., and Garman, S.C. (2012), 'The structure of human GALNS reveals the molecular basis for mucopolysaccharidosis IV A', *J Mol Biol*, 423 (5), 736-51.

Roach, P.J., Depaoli-Roach, A.A., Hurley, T.D., and Tagliabracci, V.S. (2012), 'Glycogen and its metabolism: some new developments and old themes', *Biochem J*, 441 (3), 763-87.

Roboti, P. and High, S. (2012), 'The oligosaccharyltransferase subunits OST48, DAD1 and KCP2 function as ubiquitous and selective modulators of mammalian N-glycosylation.', *J Cell Sci*, 125 (Pt 14), 3474-84.

Romijn, J.A., Coyle, E.F., Sidossis, L.S., Gastaldelli, A., Horowitz, J.F., Endert, E., and Wolfe, R.R. (1993), 'Regulation of endogenous fat and carbohydrate metabolism in relation to exercise intensity and duration', *Am J Physiol*, 265 (3 Pt 1), E380-91.

Russell, J.H. and Ley, T.J. (2002), 'Lymphocyte-mediated cytotoxicity', *Annu Rev Immunol*, 20 323-70.

Saftig, P., Hunziker, E., Wehmeyer, O., Jones, S., Boyde, A., Rommerskirch, W., Moritz, J.D., Schu, P., and von Figura, K (1998), 'Impaired osteoclastic bone resorption leads to osteopetrosis in cathepsin-K-deficient mice', *Proc Natl Acad Sci U S A*, 95 (23), 13453-58.

Saftig, P. and Klumperman, J. (2009), 'Lysosome biogenesis and lysosomal membrane proteins: trafficking meets function', *Nat Rev Mol Cell Biol*, 10 (9), 623-35.

Sagné, C., Agulhon, C., Ravassard, P., Darmon, M., Hamon, M., El Mestikawy, S., Gasnier, B., and Giros, B. (2001), 'Identification and characterization of a lysosomal transporter for small neutral amino acids', *Proc Natl Acad Sci U S A*, 98 (13), 7206-11.

Saito, S., Ohno, K., and Sakuraba, H. (2011), 'Fabry-database.org: database of the clinical phenotypes, genotypes and mutant alpha-galactosidase A structures in Fabry disease', *J Hum Genet*, 56 (6), 467-68.

Sardi, S.P., Clarke, J., Viel, C., Chan, M., Tamsett, T.J., Treleaven, C.M., Bu, J., Sweet, L., Passini, M.A., Dodge, J.C., Yu, W.H., Sidman, R.L., Cheng, S.H., and Shihabuddin, L.S. (2013), 'Augmenting CNS glucocerebrosidase activity as a therapeutic strategy for parkinsonism and other Gaucher-related synucleinopathies', *Proc Natl Acad Sci U S A*, 110 (9), 3537-42.

Sardiello, M, Palmieri, M, di Ronza, A, Medina, DL, Valenza, M, Gennarino, VA, Di Malta, C, Donaudy, F, Embrione, V, Polishchuk, RS, Banfi, S, Parenti, G, Cattaneo, E, and Ballabio, A (2009), 'A gene network regulating lysosomal biogenesis and function.', *Science*, 325 (5939), 473-77.

Sarmientos, F., Schwarzmann, G, and Sandhoff, K. (1986), 'Specificity of human glucosylceramide β-glucosidase towards synthetic glucosylsphingolipids inserted into liposomes', *Eur J Biochem*, 160 (3), 527-35.

Schapira, A.H. and Gegg, M.E. (2013), 'Glucocerebrosidase in the pathogenesis and treatment of Parkinson disease', *Proc Natl Acad Sci U S A*, 110 (9), 3214-15.

Schneede, A., Schmidt, C.K., Hölttа-Vuori, M., Heeren, J., Willenborg, M., Blanz, J., Domanskyy, M., Breiden, B., Brodesser, S., Landgrebe, J., Sandhoff, K., Ikonen, E., Saftig, P., and Eskelinen, E.L. (2011), 'Role for LAMP-2 in endosomal cholesterol transport', *J Cell Mol Med*, 15 (2), 280-95.

Schröder, B. A., Wrocklage, C., Hasilik, A., and Saftig, P. (2010), 'The proteome of lysosomes', *PROTEOMICS*, 10 (22), 4053-76.

Schulze, H and Sandhoff, K (2011), 'Lysosomal lipid storage diseases.', *Cold Spring Harb Perspect Biol*, 3 (6),

Schwake, M., Schröder, B.A., and Saftig, P. (2013), 'Lysosomal membrane proteins and their central role in physiology', *Traffic*, 14 (7), 739-48.

Scott, H.S., Anson, D.S., Orsborn, A.M., Nelson, P.V., Clements, P.R., Morris, C.P., and Hopwood, J.J. (1991), 'Human α-L-iduronidase: cDNA isolation and expression', *Proc Natl Acad Sci USA*, 88 (21), 9695-99.

Seregin, S.S and Amalfitano, A. (2011), 'Gene therapy for lysosomal storage diseases: progress, challenges and future prospects', *Curr. Pharm. Des.*, 17 (24), 2558.

Shacka, J.J., Klocke, B.J., Young, C., Shibata, M., Olney, J.W., Uchiyama, Y., Saftig, P., and Roth, K.A. (2007), 'Cathepsin D deficiency

induces persistent neurodegeneration in the absence of Bax-dependent apoptosis', *J Neurosci*, 27 (8), 2081-90.

Shayman, J.A. and Larsen, S.D. (2014), 'The Development and use of Small Molecule Inhibitors of Glycosphingolipid Metabolism for Lysosomal Storage Diseases', *J Lipid Res*,

Shearer, J and Graham, T.E. (2004), 'Novel aspects of skeletal muscle glycogen and its regulation during rest and exercise', *Exercise and sport sciences reviews*, 32 (3), 120-26.

Shental-Bechor, D. and Levy, Y. (2008), 'Effect of glycosylation on protein folding: a close look at thermodynamic stabilization', *Proc Natl Acad Sci U S A*, 105 (24), 8256-61.

Sidransky, E., Nalls, M.A., Aasly, J.O., Aharon-Peretz, J., Annesi, G., Barbosa, E.R., Bar-Shira, A., Berg, D., Bras, J., Brice, A., Chen, C.M., Clark, L.N., Condroyer, C., De Marco, E.V., Dürr, A., Eblan, M.J., Fahn, S., Farrer, M.J., Fung, H.C., Gan-Or, Z., Gasser, T., Gershoni-Baruch, R., Giladi, N., Griffith, A., Gurevich, T., Januario, C., Kropp, P., Lang, A.E., Lee-Chen, G.J., Lesage, S., Marder, K., Mata, I.F., Mirelman, A., Mitsui, J., Mizuta, I., Nicoletti, G., Oliveira, C., Ottman, R., Orr-Urtreger, A., Pereira, L.V., Quattrone, A., Rogaeva, E., Rolfs, A., Rosenbaum, H., Rozenberg, R., Samii, A., Samaddar, T., Schulte, C., Sharma, M., Singleton, A., Spitz, M., Tan, E.K., Tayebi, N., Toda, T., Troiano, A.R., Tsuji, S., Wittstock, M., Wolfsberg, T.G., Wu, Y.R., Zabetian, C.P., Zhao, Y., and Ziegler, S.G. (2009), 'Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease', *N Engl J Med*, 361 (17), 1651-61.

Sim, L., Quezada-Calvillo, R., Sterchi, E.E., Nichols, B.L., and Rose, D.R. (2008), 'Human intestinal maltase-glucoamylase: crystal structure of the N-terminal catalytic subunit and basis of inhibition and substrate specificity', *J Mol Biol*, 375 (3), 782-92
.
Sim, L., Jayakanthan, K., Mohan, S., Nasi, R., Johnston, B.D., Pinto, B.M., and Rose, D.R. (2010a), 'New glucosidase inhibitors from an ayurvedic herbal treatment for type 2 diabetes: structures and inhibition of human intestinal maltase-glucoamylase with compounds from *Salacia reticulata*', *Biochemistry*, 49 (3), 443-51.

Sim, L., Willemsma, C., Mohan, S., Naim, H.Y., Pinto, B.M., and Rose, D.R. (2010b), 'Structural basis for substrate selectivity in human maltase-glucoamylase and sucrase-isomaltase N-terminal domains', *J Biol Chem*, 285 (23), 17763-70.

Simon, S.M. and Blobel, G. (1991), 'A protein-conducting channel in the endoplasmic reticulum', *Cell*, 65 (3), 371-80.

Sleat, D.E., Lackland, H., Wang, Y., Sohar, I., Xiao, G., Li, H., and Lobel, P. (2005), 'The human brain mannose 6-phosphate glycoproteome: a complex mixture composed of multiple isoforms of many soluble lysosomal proteins', *PROTEOMICS*, 5 (6), 1520-32.

Sleat, D.E., Della Valle, M.C., Zheng, H., Moore, D.F., and Lobel, P. (2008), 'The mannose 6-phosphate glycoprotein proteome', *J Proteome Res*, 7 (7), 3010-21.

Sleat, DE, Ding, L, Wang, S, Zhao, C, Wang, Y, Xin, W, Zheng, H, Moore, DF, Sims, KB, and Lobel, P (2009), 'Mass spectrometry-based protein profiling to determine the cause of lysosomal storage diseases of unknown etiology.', *Mol Cell Proteomics*, 8 (7), 1708-18.

Soubannier, V., McLelland, G.L., Zunino, R., Braschi, E., Rippstein, P., Fon, E.A., and McBride, H.M. (2012), 'A vesicular transport pathway shuttles cargo from mitochondria to lysosomes', *Curr Biol*, 22 (2), 135-41.

Sousa, M. and Parodi, A.J. (1995), 'The molecular basis for the recognition of misfolded glycoproteins by the UDP-Glc: glycoprotein glucosyltransferase', *EMBO J*, 14 (17), 4196.

Steet, R., Lee, W. S., and Kornfeld, S. (2005), 'Identification of the minimal lysosomal enzyme recognition domain in cathepsin D', *J Biol Chem*, 280 (39), 33318-23.

Stern, R. and Jedrzejas, M.J. (2006), 'Hyaluronidases: their genomics, structures, and mechanisms of action', *Chem Rev*, 106 (3), 818-39.

Straus, Werner (1958), 'Colorimetric analysis with N, N-dimethyl-p-phenylenediamine of the uptake of intravenously injected horseradish peroxidase by various tissues of the rat', *J Biophys and -Biochem cyt*, 4 (5), 541-50.

Suter, A., Everts, V., Boyde, A., Jones, S.J., Lüllmann-Rauch, R., Hartmann, D., Hayman, A.R., Cox, T.M., Evans, M.J., Meister, T., von Figura, K., and Saftig, P. (2001), 'Overlapping functions of lysosomal acid phosphatase (LAP) and tartrate-resistant acid phosphatase (Acp5) revealed by doubly deficient mice', *Development*, 128 (23), 4899-910.

Szilágyi, A. and Závodszky, P. (2000), 'Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey', *Structure*, 8 (5), 493-504.

Taka, J., Ogasahara, K., Jeyakanthan, J., Kunishima, N., Kuroishi, C., Sugahara, M., Yokoyama, S., and Yutani, K. (2005), 'Stabilization due to dimer formation of phosphoribosyl anthranilate isomerase from *Thermus thermophilus* HB8: X-ray analysis and DSC experiments', *J Biochem*, 137 (5), 569-78.

Tayebi, N., Walker, J., Stubblefield, B., Orvisky, E., LaMarca, M.E., Wong, K., Rosenbaum, H., Schiffmann, R., Bembi, B., and Sidransky, E. (2003), 'Gaucher disease with parkinsonian manifestations: does glucocerebrosidase deficiency contribute to a vulnerability to parkinsonism?', *Mol Genet Metab*, 79 (2), 104-9.

Tiede, S., Storch, S., Lübke, T., Henrissat, B., Bargal, R., Raas-Rothschild, A., and Braulke, T. (2005), 'Mucolipidosis II is caused by mutations in GNPTA encoding the α/β GlcNAc-1-phosphotransferase.', *Nat Med*, 11 (10), 1109-12.

Tikkanen, R., Enomaa, N., Riikonen, A., Ikonen, E., and Peltonen, L. (1995), 'Intracellular sorting of aspartylglucosaminidase: the role of N-linked oligosaccharides and evidence of Man-6-P-independent lysosomal targeting', *DNA Cell Biol*, 14 (4), 305-12.

Tikkanen, R., Peltola, M., Oinonen, C., Rouvinen, J., and Peltonen, L. (1997), 'Several cooperating binding sites mediate the interaction of a lysosomal enzyme with phosphotransferase', *EMBO J*, 16 (22), 6684-93.

Tollersrud, O.K., Berg, T., Healy, P., Evjen, G., Ramachandran, U., and Nilssen, O. (1997), 'Purification of bovine lysosomal alpha-mannosidase, characterization of its gene and determination of two mutations that cause alpha-mannosidosis', *Eur J Biochem*, 246 (2), 410-19.

Toscano, A., Montagnese, F., and Musumeci, O. (2013), 'Early is better? A new algorithm for early diagnosis in late onset Pompe disease (LOPD)', *Acta Myol*, 32(2) (2), 78-81.

Trombetta, E.S. and Parodi, A.J. (2003), 'Quality control and protein folding in the secretory pathway', *Annu Rev Cell Dev Biol*, 19 649-76.

Tsuji, A., Omura, K., and Suzuki, Y. (1988), 'I-cell disease: evidence for a mannose 6-phosphate independent pathway for translocation of lysosomal enzymes in lymphoblastoid cells', *Clin Chim Acta*, 176 (1), 115-21.

Turk, D., Janjić, V., Štern, I., Podobnik, M., Lamba, D., Dahl, S.W., Lauritzen, C., Pedersen, J., Turk, V., and Turk, B. (2001), 'Structure of human dipeptidyl peptidase I (cathepsin C): exclusion domain added to an endopeptidase framework creates the machine for activation of granular serine proteases', *EMBO J*, 20 (23), 6570-82.

Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B., and Turk, D. (2012), 'Cysteine cathepsins: from structure, function and regulation to new frontiers', *Biochim Biophys Acta*, 1824 (1), 68-88.

Valayannopoulos, V. and Wijburg, F.A. (2011), 'Therapy for the mucopolysaccharidoses', *Rheumatology*, 50 Suppl 5 v49-59.

van den Hout, H.M., Hop, W., van Diggelen, O.P., Smeitink, J.A., Smit, G.P., Poll-The, B.T., Bakker, H.D., Loonen, M.C., de Klerk, J.B., Reuser, A.J., and van der Ploeg, A.T. (2003), 'The natural course of infantile Pompe's disease: 20 original cases compared with 133 cases from the literature', *Pediatrics*, 112 (2), 332-40.

van den Hout, J.M., Reuser, A.J., de Klerk, J.B., Arts, W.F., Smeitink, J.A., and Van der Ploeg, A.T. (2001), 'Enzyme therapy for pompe disease with recombinant human α-glucosidase from rabbit milk.', *J Inherit Metab Dis*, 24 (2), 266-74.

van der Ploeg, A.T. and Reuser, A.J. (2008), 'Pompe's disease', *Lancet*, 372 (9646), 1342-53.

van der Ploeg, A.T., Clemens, P.R., Corzo, D., Escolar, D.M., Florence, J., Groeneveld, G.J., Herson, S., Kishnani, P.S., Laforet, P., Lake, S.L., Lange, D.J., Leshner, R.T., Mayhew, J.E., Morgan, C., Nozaki, K., Park, D.J., Pestronk, A., Rosenbloom, B., Skrinar, A., van Capelle, C.I., van der Beek, N.A., Wasserstein, M., and Zivkovic, S.A. (2010), 'A randomized study of alglucosidase alfa in late-onset Pompe's disease', *N Engl J Med*, 362 (15), 1396-406.

van der Spoel, A., Bonten, E., and d'Azzo, A. (1998), 'Transport of human lysosomal neuraminidase to mature lysosomes requires protective protein/cathepsin A', *EMBO J*, 17 (6), 1588-97.

Van Lenten, B.J. and Fogelman, A.M. (1990), 'Processing of lipoproteins in human monocyte-macrophages', *J Lipid Res*, 31 (8), 1455-66.

Vanto, T., Salmi, T.T., Kalimo, H., Lang, H., Näntö, V., Berlin, M., and Penttinen, R. (1982), '[Pompe's disease or glycogen storage disease].', *Duodecim*, 98 (9), 709-16.

Varki, A. and Kornfeld, S. (1980), 'Structural studies of phosphorylated high mannose-type oligosaccharides', *J Biol Chem*, 255 (22), 10847-58.

Vega, M.A., Rodriguez, F., Segui, B., Calés, C., Alcalde, J., and Sandoval, I.V. (1991), 'Targeting of lysosomal integral membrane protein LIMP II. The tyrosine-lacking carboxyl cytoplasmic tail of LIMP II is sufficient for direct targeting to lysosomes.', *Journal of Biological Chemistry*, 266 (25), 16269-72.

Venditti, R., Wilson, C., and De Matteis, M.A. (2014), 'Exiting the ER: what we know and what we don't', *Trends Cell Biol*, 24 (1), 9-18.

Vitale, N., Caumont, A.S., Chasserot-Golaz, S., Du, G., Wu, S., Sciorra, V.A., Morris, A.J., Frohman, M.A., and Bader, M.F. (2001), 'Phospholipase D1: a key factor for the exocytotic machinery in neuroendocrine cells', *EMBO J*, 20 (10), 2424-34.

von Bülow, R., Schmidt, B., Dierks, T., von Figura, K., and Usón, I. (2001), 'Crystal structure of an enzyme-substrate complex provides insight into the interaction between human arylsulfatase A and its substrates during catalysis', *J Mol Biol*, 305 (2), 269-77.

Walkley, S.U. (2009), 'Pathogenic cascades in lysosomal disease - Why so complex?', *J Inherit Metab Dis*, 32 (2), 181-89.

Warner, J.B., Thalhauser, C., Tao, K., and Sahagian, G.G. (2002), 'Role of *N*-linked oligosaccharide flexibility in mannose phosphorylation of lysosomal enzyme cathepsin L', *J Biol Chem*, 277 (44), 41897-905.

Wiedmann, M., Kurzchalia, T.V., Hartmann, E., and Rapoport, T.A. (1987), 'A signal sequence receptor in the endoplasmic reticulum membrane', *Nature*, 328 (6133), 830-33.

Wilke, S., Krausze, J., and Büssow, K. (2012), 'Crystal structure of the conserved domain of the DC lysosomal associated membrane protein: implications for the lysosomal glycocalyx', *BMC Biol*, 10 62.

Willemsen, R, van der Ploeg, AT, Busch, HF, Zondervan, PE, Van Noorden, CJ, and Reuser, AJ (1993), 'Synthesis and in situ localization of lysosomal alpha-glucosidase in muscle of an unusual variant of glycogen storage disease type II', *Ultrastruct Pathol*, 17 (5), 515-27.

Winchester, B. (2005), 'Lysosomal metabolism of glycoproteins', *Glycobiology*, 15 (6), 1R-15R.

Winchester, B. (2012), 'Laboratory Diagnosis of Lysosomal Storage Diseases', (Lysosomal Storage Disorders: A Practical Guide, John Wiley & Sons, Ltd), 20-28.

Wisselaar, H.A., Kroos, M.A., Hermans, M.M., van Beeumen, J., and Reuser, A.J. (1993), 'Structural and functional changes of lysosomal acid α-glucosidase during intracellular transport and maturation', *J Biol Chem*, 268 (3), 2223-31.

Withana, N.P., Blum, G., Sameni, M., Slaney, C., Anbalagan, A., Olive, M.B., Bidwell, B.N., Edgington, L., Wang, L., Moin, K., Sloane, B.F., Anderson, R.L., Bogyo, M.S., and Parker, B.S. (2012), 'Cathepsin B inhibition limits bone metastasis in breast cancer', *Cancer Res*, 72 (5), 1199-209.

Wormald, M.R. and Dwek, R.A. (1999), 'Glycoproteins: glycan presentation and protein-fold stability', *Structure*, 7 (7), R155-60.

Wraith, E.J., Hopwood, J.J., Fuller, M., and Meikle, P.J. (2005), 'Laronidase treatment of mucopolysaccharidosis I', *BioDrugs*, 19 (1), 1-7.

Wright, C.S., Li, S.C., and Rastinejad, F. (2000), 'Crystal structure of human GM2-activator protein with a novel β-cup topology', *J Mol Biol*, 304 (3), 411-22.

Xu, S., Zhao, L., Larsson, A., and Venge, P. (2009), 'The identification of a phospholipase B precursor in human neutrophils', *FEBS J*, 276 (1), 175-86.

Yaghootfam, A., Schestag, F., Dierks, T., and Gieselmann, V. (2003), 'Recognition of arylsulfatase A and B by the UDP-*N*-acetylglucosamine: lysosomal enzyme *N*-acetylglucosamine-phosphotransferase', *J Biol Chem*, 278 (35), 32653-61.

Yang, C.F., Liu, H.C., Hsu, T.R., Tsai, F.C., Chiang, S.F., Chiang, C.C., Ho, H.C., Lai, C.J., Yang, T.F., Chuang, S.Y., Lin, C.Y., and Niu, D.M. (2014), 'A large-scale nationwide newborn screening program for Pompe

disease in Taiwan: towards effective diagnosis and treatment', *Am J Med Genet A*, 164A (1), 54-61.

Yoshimizu, M., Tajima, Y., Matsuzawa, F., Aikawa, S., Iwamoto, K., Kobayashi, T., Edmunds, T., Fujishima, K., Tsuji, D., Itoh, K., Ikekita, M., Kawashima, I., Sugawara, K., Ohyanagi, N., Suzuki, T., Togawa, T., Ohno, K., and Sakuraba, H. (2008), 'Binding parameters and thermodynamics of the interaction of imino sugars with a recombinant human acid α-glucosidase (alglucosidase alfa): insight into the complex formation mechanism', *Clin Chim Acta*, 391 (1-2), 68-73.

Zachos, C., Blanz, J., Saftig, P., and Schwake, M. (2012), 'A critical histidine residue within LIMP-2 mediates pH sensitive binding to its ligand β-glucocerebrosidase.', *Traffic*, 13 (8), 1113-23.

Zhao, J., Brault, J.J., Schild, A., Cao, P., Sandri, M, Schiaffino, S, Lecker, SH, and Goldberg, AL (2007), 'FoxO3 coordinately activates protein degradation by the autophagic/lysosomal and proteasomal pathways in atrophying muscle cells', *Cell Metab*, 6 (6), 472-83.

Zhou, M., Ma, Z., and Sly, W.S. (1995), 'Cloning and expression of the cDNA of chicken cation-independent mannose-6-phosphate receptor.', *Proc Natl Acad Sci U S A*, 92 (21), 9762-66.

Zielinska, D.F., Gnad, F., Wiśniewski, J.R., and Mann, M. (2010), 'Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints', *Cell*, 141 (5), 897-907.